AD 668847

Best Available Copy

PROJECT INTREX

Massachusetts Institute of Technology
Cambridge, Massachusetts

15 March 1968

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT INTREX

SEMIANNUAL ACTIVITY REPORT

15 September 1967 to 15 March 1968

PR-5

15 March 1968

CAMBRIDGE MASSACHUSETTS

ACKNOWLEDGEMENTS

The research reported in this document was made possible through the support extended the Massachusetts Institute of Technology, Project Intrex, under a grant from the Carnegie Corporation, under Contract NSF-C472 from the National Science Foundation and the Advanced Research Projects Agency of the Department of Defense, and under a grant from the Council on Library Resources, Inc.

TABLE OF CONTENTS

PROJECT INTREX

Activity Report

I. INTRODUCTION

"Endless Volumes" is the title of an editorial in which the Times Literary Supplement[1] of London laments the proliferation of our scientific literature. After citing some frightening statistics on the annual production of journal articles, the editorial suggests that the individual scientific paper may have become "an unnecessary and undesirable luxury", and that we might be better off with integrated articles written by knowledgeable reviewers. In commenting upon this suggestion in a subsequent Letter to the Editor,[2] B. C. Brookes, of University College, London, refers to Project Intrex and credits us with the proposal "to dispense with conventional journals and books by publishing 'journals' electronically within an 'online' computerized information transfer network which offers immediate sight of any published paper within the network to the scientist at his desk, and which offers printouts on palpable paper only to those backwoodsmen who have no access to the network".

Now this important notion of the "publishing" potential of an "online" computer community was indeed discussed at the Intrex Planning Conference of 1965. But it was only one of many ideas discussed in that stimulating five weeks' discourse. It was never considered as a complete answer or as the only answer to the problem of the growth of scientific literature. The consensus was very clear that a complete solution for libraries must be sought at the confluence of three streams: The modernization of current library procedures; the growth of national information networks; and the extension of interactive computer communities into the domain of the library.

Another common misconception about Intrex is that the project is a plan for the earliest possible conversion of the entire holdings of the M.I.T. Libraries (approx. one million volumes) to computer storage. Of course, we are quite optimistic about the prospects for a catalog that is digitally-encoded and computer-manipulated, and we are at work on an experimental model of such a catalog. We are also working on the problems of displaying the full text of documents at locations remote from the library. Those problems are much easier for digitally-encoded text than for graphic records, and we therefore hope that libraries will come to store some of their materials in digital form in computer memories. But we know that economical mass memories are not yet within our reach, and we are devoting substantial efforts to the

---

[1] 28 December 1967

[2] Times Literary Supplement, 11 January 1968

problems of handling graphic records in information networks. We do not expect to see the early extinction of the Gutenberg galaxy, even within the limited confines of our own university.

After countering these misunderstandings with explanations of what Project Intrex is not, it may be well to state once again what it really is. It is a program of experiments intended to provide a foundation for the design of future information transfer systems. We visualize the library of the future as a computer-managed communications network, but we do not know today how to design such a network in all its detail. We lack the necessary experimental facts, especially in the area of the user's interaction with the system. We want to discover these facts by experimentation not only in the laboratory but above all in the real-life environment of an operating library.

We have concentrated our initial efforts on the problems of access--bibliographic access through an augmented catalog, and access to full text. In the M.I.T. Electronic Systems Laboratory, under the direction of Professor Reintjes, a steadily growing team of faculty, students, and research staff is engaged in the exploration of these two initial problems. This report describes their activities during the past six months.

Carl F. J. Overhage

Cambridge, Massachusetts

15 March 1968

II. RESEARCH AND DEVELOPMENT ACTIVITIES (Electronic Systems Laboratory)

A. STATUS OF THE PROGRAM

Professor J. F. Reintjes

Our efforts during the past six months have concentrated on the implementation of our experimental library information storage and retrieval system. The configuration of the initial system is defined in detail and our objective now is to bring it to a state in which it can be tested and evaluated as a complete system.

It will be recalled from our preceding reports that two forms of storage media are being employed--computer disk-file storage and photographic film. Our augmented catalog of at least 10,000 journal articles, reports and theses in selected areas of materials science and engineering will be contained in the disk files of the M.I.T.-modified 7094 multi-access computer, and the full text of the same 10,000 items, including all pictorial information, will be stored on microfiche.

The retrieval mechanisms consist of the library user working with the multi-access computer in an online mode in accordance with procedures governed by our retrieval programs, and an automatic microfiche retrieval and image-transmission system. The multi-access computer is engaged in order to identify material in the catalog which may be pertinent to a user's requirements, and the image-transmission system enables him to obtain, at the location of his computer terminal, the full text of the information being sought.

In order to determine the most desirable features of a library-information-type computer console for use with the augmented catalog, we are developing an experimental console with flexible characteristics. Our ability to present a console with a variety of features to a community of users should enable us, ultimately, to arrive at a consensus of preferred attributes.

Details of progress being made on the development of the augmented catalog, the augmented-catalog display console, the storage and retrieval programs, and the full-text-access system are presented in the sections that follow.

B. THE COMPUTER-STORED AUGMENTED-CATALOG PROGRAM

1. AUGMENTED-CATALOG INPUTTING

Staff Members

Mr. A. R. Benenfeld
Mrs. S. K. Escudier
Mrs. E. J. Gurley
Mrs. S. F. Lage
Miss L. T. Lee
Miss S. P. Niessen
Professor J. F. Reintjes
Miss J. E. Rust

Cataloger Assistants

Mrs. D. Bell
Mrs. A. M. Davis
Mrs. C. U. Koss

Graduate Students

Mr. U. Chinwah
Mr. N. A. Clark

Undergraduate Students

Miss M. D. Beaudry
Mr. S. C. Chamberlain
Miss C. Daniel
Mr. L. A. Distaso
Mr. R. R. Doering
Mr. H. D. Feldman
Mr. J. F. Kaar
Mr. M. D. Katz
Mr. R. M. Koolish
Mr. R. C. Lufkin
Mr. G. H. McKinney
Mr. S. M. Neirman
Mr. C. T. Pynn
Mr. J. S. Rothman
Mr. T. F. Wagner

SUMMARY

Literature in a second materials science research area, high-temperature metallurgy, is being added to the data base. Changes in document-selection procedures have been made in an effort to ensure a highly relevant catalog. An authority file of corporate names and explicit structuring of such names have been established. Changes in the input workflow have been made such that: a cataloger-assistant now handles most of the descriptive cataloging; only one generation of punched paper tape is produced; and editing of proofread computer printout copy is handled by online techniques. A successful student indexing program has been established. As of March 1, 1968,3240 documents have been indexed.

DATA BASE AND LITERATURE SELECTION

The augmented-catalog data base is currently being built upon the literature for two research areas in materials science and engineering. These two areas and their subdivisions are:

A. Radio-frequency, microwave, and optical spectroscopy of liquids and solids

1. Magnetic properties of materials

2. Critical phenomena and phase transitions

3. Use of light scattering in the study of properties of materials

4. Use of ultrasonics in the study of properties of materials

B. High-temperature metallurgy

1. Dispersed-particle strengthening

2. Creep rupture properties

3. Fatigue behavior

4. Oxidation resistance

5. Mechanisms of strengthening

6. Alloy phase studies of high-temperature materials

It should be noted that these particular subdivisions reflect the interests of M.I.T. research groups who are active in the two principal areas. At the time of writing this report, the concluding phase of an analysis is underway which will add two or three new materials-science research-literature areas to the data base.

An analysis of the journal-article literature selected by librarians for research-area A indicated that much peripheral material was being included. For example, in one test, the librarians selected 186 of 585 possible articles (31.8 percent), whereas a professor and a doctoral candidate in research area A selected 49 (8.4 percent) and 66 (11.3 percent) articles, respectively, from the same group of 585. Of those articles selected by the professor and doctoral candidate, the librarians had selected 73.5 percent and 63.6 percent, respectively, while the doctoral candidate selected only 46.9 percent of the articles selected by the professor. Apparently, the librarian's selection of a large proportion of relevant articles is achieved through the scatter effect of overselection. Accordingly, in order not to overextend the data base, a change was made in the selection procedure to provide more direct feedback on the appropriateness of individual documents for the collection. Two participants in research-area A, Professor J. Litster and Mr. N. Clark, agreed to indicate from journal-issue tables of contents routed to them, articles they consider important to their area. It is expected that a similar procedure will be followed for the other research areas.

Additionally, Mr. Clark is providing valuable service through his ability to assist the librarians on matters relating to his field. He is also aiding us in the development of quality controls in the indexing operation.

DATA ELEMENTS AND CODING

To maintain consistency in recording corporate names, whether as an author or as an author affiliation, an authority file of corporate names has been established.

This file is currently on cards. Procedures for machine-storage of the file and the referencing of established corporate names by a number associated with the name are being investigated.

The structure for an established corporate name has been made explicit. A corporate name consists of a main heading which may have one or more subheadings; two spaces separate each subheading. Whenever possible, the geographic location pertaining to the name given by the last subheading is added to the entry, the addition being made to the main heading; an exception occurs when the place is already part of the main heading. All place names, whether they appear within the main heading or as an addition to it, are tagged by enclosing the place name, or its established abbreviation, within slanted single quotes. Names that are qualified by the addition of a larger associated corporate body name have that addition enclosed within square brackets, and that addition may be further qualified by place. Conference names are separately tagged by preceding the name with a sharp symbol (#). Examples of these corporate name structures are:

> Northwestern University, `Evanston´, `Ill.´ Materials Research Center.
>
> University of `California´, `Berkeley´. Electronics Research Laboratory.
>
> `Swarthmore´ College, (`Pa.´) Dept. of Physics.
>
> Homer Research Laboratories [`Bethlehem´ Steel Co., (`Pa.´)]
>
> #Conference on Magnetism and Magnetic Materials, 12th, `Washington´, `D.C.´, Nov. 15 1966.

Access Number has been established as Field 5. This new field contains the microfiche number, and the beginning-frame and end-frame numbers of the microfiche text of the document described in a record. These machine-stored numbers serve to couple the text-access experiments to the augmented-catalog experiments. The numbers are established at the time of microfilming a document and are recorded by the operator onto the microfilm-initiation slip, a copy of which is returned with a Xerox print of the document to the catalog input group. Otherwise, the microfilm-initiation procedure remains essentially unchanged from the description given in the preceding Semiannual Activity Report. New procedures in the film and fiche preparation processes are described in Section C of the present report.

Several minor changes have been made to the Cataloging Manual. Preparation of a new edition of this manual is now being considered.

PROCESSING

Several changes have been made in the workflow. A block diagram of the present inputting process, shown in Fig. 1, reflects these changes.

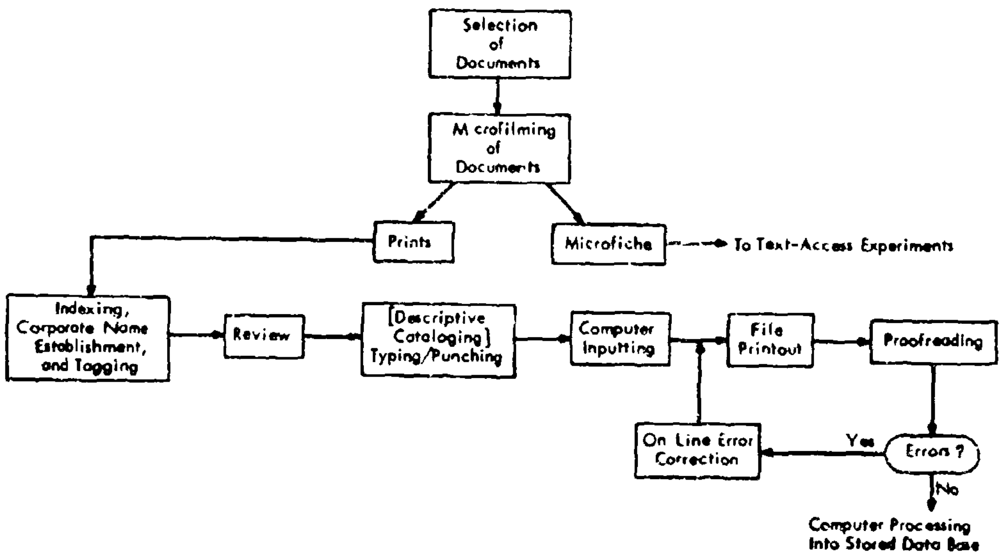


Fig. 1 Present Workflow of the Augmented-Catalog Inputting Process

The preceding Semiannual Activity Report indicated the responsibility for gathering and formatting descriptive cataloging data was being shifted to the typists. This change has been successful. Essentially, the only descriptive cataloging of journal articles now being done by the catalogers is the establishment of corporate-author names, listing special features of the document, and tagging for the typist's attention information for only a few other fields (for example, excerpts or inclusion of a table of contents).

In the typing operation, ten records are now batched to form one file. This procedure has considerably reduced the amount of handling required in the input of the punched paper tape to the computer and it has simplified the interim handling of records during the proofreading and error-correction operations.

The most significant changes in the processing are the use of computer printouts for the proofreading and the correction of errors through an online-editing technique. This means that only one generation of punched tape is now prepared. Additionally, online editing allows dialog between typist and computer for immediate verification of changes in the data so that reintroduction of errors (as might be the case in preparing successive generations of punched tapes) is potentially eliminated. Proofreading of the first printout of a file is performed by a cataloger; this is also the first time that the descriptive cataloging performed by the typists is checked. A typist does the online editing of files and she also proofreads successive generations of file printouts. During the first proofreading of each file, a tabulation is made of the number and class of errors (cataloging, policy, typing, mechanical); the online editing typist makes a similar tabulation for additional errors not previously spotted. When no further errors are detected in a file, the online-editing program is used to certify that the file is ready for processing into the computer-stored data base. Further details on the online editing procedure and preliminary data for its operation are discussed in the next section.

As of March 1, 1968,

3240	documents have been indexed,
2970	records have been reviewed,
2730	records have been typed,
266	files have entered the computer and codes converted from Flexowriter to ASCII codes,
218	files have passed through the first proofreading and editing process,
171	files have passed through a second proofreading and editing process,
91	files have passed through three or more proofreading and editing processes,
170	files have been certified for further processing into the stored data base.

INDEXING

The indexing process requires the greatest amount of the professional effort expended in cataloging. Our current procedures call for the use of terms (which generally are combinations of noun phrases) based upon the text of a document. A term may require further intensification to provide fairly complete context among its individual components. Each term is structured to provide sufficient expression of a concept such that the term may stand by itself. Further, each term is weighted to reflect that proportion of a document devoted to discussing the represented concept. There is no limit on the number of terms assigned to a document and no authority list of terms is used. This is not a rapid indexing technique. Since any retrieval system depends upon the indexing base, this technique does allow for flexibility in the designs of experimental information retrieval systems. However, in order to increase the rate of cataloging without sacrificing the basis of the indexing procedure, a student indexing program was instituted last November.

Students, predominantly undergraduates in science and engineering, are now being employed to index documents. Prior to assuming indexing duties, these students attend three training sessions and receive practice indexing assignments for homework following the first two sessions. The first session, lasting one and one-half hours, (1) orients them to the experimental work of the project, (2) covers the general conditions under which they will work, (3) covers the specific routines of the job, (4) introduces the nature of the indexing and weighting process with suggestions on how to proceed, and (5) reviews a specific example of a previously indexed document. The homework assignment for each student consists of the same set of five previously indexed documents, together with the full cataloging records for those documents. The students are asked to index the documents, to compare their work with the cataloger's work, and to study the derivation of the terms and weights.

The second group session is devoted to considerations of the indexing procedure based upon the practice set of five documents. At this session the indexing of these articles by other staff personnel is presented so that the student can see a spectrum of acceptable indexing. The students are encouraged to develop their own acceptable and consistent approach to indexing. The second homework assignment consists of indexing another set of five documents. This time the assigned documents have not been previously indexed, and each student has a different set. The student's homework is commented upon and returned to him at the third class session. The review at this stage is very detailed because the students are essentially indexing completely on their own for the first time. The last session is a combination of individual meetings with the students to explain the comments on their work and a group session to convey, at large, experiences and comments.

At the last group session, a student is assigned to a librarian who will serve as his reviewer. No more than two students are assigned to a librarian. Further training and guidance of the student in indexing is provided on the job by the reviewer. Although encouraged to do their work at the laboratory, the students may work at home; each student is expected to work about ten hours a week.

The cataloging for which each student is responsible is the indexing of a document (field 73), the author's purpose (field 65), and the level of approach (field 66). The student's reviewer critiques the indexing more thoroughly than would be done for another librarian's indexing. Corrections, additions, or deletions to the indexing are analyzed and these are discussed with the student, generally once a week. The reviewer also adds other necessary fields (for example, established corporate names) to the student-initiated record before it enters the normal work-flow pattern at the typing stage.

Our first experiences in using students as indexers have been varied, but overall, we consider the program to be successful. Of the ten students who began in the initial program, six have shown the ability to produce quite acceptable indexing work. The rapport between student and reviewer has generally been very good. Each provides the other with a better insight and perspective of the nature of the indexing process. Five students of our original group of ten have terminated either because of lack of time to work or because of unacceptable work. One difficulty that the students encounter is that their available working time is influenced by the rigors of their academic programs and examination periods. This factor, in turn, influences the rate at which cataloging can progress with this kind of assistance. With the start of the Spring semester, additional students have been recruited to bring our complement to twelve.

EVALUATION

During the Spring 1968 semester, Mr. Richard Lufkin, an undergraduate student at M.I.T., as part of his S.B. thesis requirement, will undertake an analysis of the learning curves of indexers as a function of time since beginning employment and as a function of the subject area of the documents cataloged. This study is an extension of an earlier analysis made immediately after our cataloging began.

A procedure to determine the degree to which an indexer's terms cover the content of a document is under development. Continuing attention is being given to new approaches to subject indexing which will optimize the extent of document content covered by context-indexing, the length and number of index terms, the indexing quality, and the indexing time. Strengthening the data encoding is also a continuing effort.

Test experiments for online cataloging, and for offline cataloging but with online inputting, which would be compared and evaluated with present procedures, are in the planning stage.

2. STORAGE AND RETRIEVAL

Staff Members

Mr. C. E. Hurlburt
Mr. P. Kugel
Mr. R. L. Kusik
Mr. R. S. Marcus
Mr. M. K. Molnar
Professor J. F. Reintjes
Professor A. K. Susskind
Mr. H. F. Vandevenne (visiting)

Graduate Students

Mr. R. Goldschmidt
Mr. F. Guertin
Mr. W. Kampe
Mr. T. Welch

Undergraduate Students

Mr. R. Greer
Mr. T. Lin
Mr. K. Pogran
Mr. R. Voss

SUMMARY

As described in the preceding Semiannual Activity Report, the implementation of our experimental augmented-catalog storage-and-retrieval system has been scheduled in a research program of three phases:

Phase I: A restricted, basic system for use by the Intrex staff in testing and evaluating various techniques of file organization, storage, and retrieval.

Phase II: A more complete system with which we plan to conduct experiments on storage and retrieval in the context of our 10,000-document augmented catalog and a selected group of users. The present M.I.T.-modified IBM-7094 time-sharing computing system will be used for these experiments.

Phase III: An expanded version of the Phase-II system. It is expected that Phase III will be implemented on the next-generation time-sharing computers at M.I.T. which will be coming into operation in late 1968.

The Phase-I storage-and-retrieval system is now operational and is serving its intended purpose as an analysis tool. An initial version of the Phase-II system has been designed and parts of it have been implemented. This initial version of Phase II is scheduled to be operational in spring, 1968. We have continued to study various topics pertaining to the Phase-III system which would be operational later in 1968. These topics include automatic indexing and thesaurus generation.

In addition to the above systems, which are intended as a basis for experiments in the near future, a study of computer-system organization for a much larger catalog is going forward. A capacity to store an augmented catalog for one million documents is presupposed in this study. Two aspects of this study investigated during this reporting period are: large-capacity digital-storage devices having the features of low cost, rapid access, and high reliability; and effective organization of large files.

THE PHASE-I SYSTEM

General

The Phase-I storage-and-retrieval system described in the preceding Semiannual Activity Report has been made operational and is being used to test how well certain of our file organization, storage, and retrieval techniques work on the present M.I.T.-modified IBM-7094 based time-sharing computing system (CTSS). Thus far, based on files for sample batches of catalog records consisting of several hundred records, these tests indicate that our techniques should be satisfactory for our Phase-II experimentation. Observed time values for some of the critical operations are given below. It should be pointed out that these values are a consequence of the particular software-hardware combination we are presently using in the CTSS system and do not represent optimum values that are possible for magnetic-disk hardware. (In particular, note the analysis for a disk-oriented large augmented catalog given in the preceding Semiannual Activity Report.)

Processing of Catalog Paper Tapes

The inputting procedure for converting catalog information stored on paper tape to ASCII-coded magnetic form described in the preceding Semiannual Activity Report is operating quite satisfactorily. Mr. Kenneth Pogran has written system utility programs which enable a more automatic handling of the data. Preliminary results indicate that an average of 6.2 seconds of 7094 CTSS time are required to process one catalog record to the point where it is ready for editing. This time corresponds to a cost of approximately 50 cents per record.

The use of an online context-editing program (see below) has enabled us to eliminate line numbering of the printouts and thus to reduce the cost of the production of printouts by about 80 percent. This is a meaningful saving since a new printout is made for each iteration of the editing process (a catalog record usually undergoes two or three iterations in this process). Efforts are continuing to make this production task even more efficient.

Online Editing

The editing of catalog records is now being done online by means of a context-editing program. The procedure is illustrated in Fig. 2. Once a batch of ten catalog records has been placed in the computer, it is converted into ASCII code, stored on the disk, and printed on a 1403 line printer equipped with an extended character-set print chain. This printout is then proofread and errors marked. Since the proofreading is performed from the printouts, it is not tied online to the computer. After the errors have been identified, they are corrected online using an IBM 2741 console. This is accomplished through use of a context editing program which enables the typist to identify the change to be made by its context and to see immediately the results of her instructions to the computer.

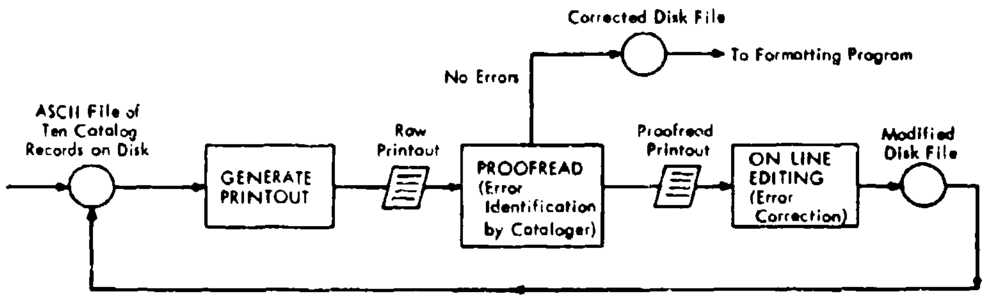


Fig. 2 The Editing Procedure

The numbers in the following discussion refer to line numbers of Fig. 3 which illustrates a typical online-editing dialog. The typist first specifies the line

```
/within the framework/                                                          (1)
        within the framework of the simple convering collision-time model.  It is shown (2)
v/convering/conserving/                                                         (3)
        within the framework of the simple convering collision-time model.  It is shown (4)
s                                                                               (5)
```

Fig. 3 Sample Online Editing Dialog

to be corrected by typing a suitably unique portion of that line. (See line 1.) The computer finds a line with that specification and responds (line 2) by typing the entire line so that the typist can determine if it is indeed the desired line. Assuming that it was, she then indicates which characters are to be changed, and in what manner (line 3). Thus, in the example, the word <u>conserving</u> has been misspelled and is to be corrected. (The command, v, indicates that she wants to verify the substitution before it is actually made.) The computer, therefore, retypes the entire line, emphasizing those letters which it understood are to be corrected by typing them in red, signified in our figure by a double underline (line 4). If this is indeed the correct substitution, the typist types "s" (substitute) (line 5) and the change is made. If her specification is ambiguous, the computer finds each ambiguity, within the correct line, and presents it as described above. When the intended characters are typed in red, the typist can give the <u>s</u> (substitution) command. For the other cases she simply types a carriage return and no action is taken. In addition to making substitutions, the typist can also delete lines or insert new lines.

After this online editing a new printout is generated (see Fig. 2). When the proofreader is satisfied that the file is correct, the batch leaves the editing loop for processing into the data base.

Preliminary data from a sample of 320 catalog records through one iteration of the editing loop (a catalog record usually undergoes two or three iterations in this loop) indicate that it takes about three seconds of computer time per catalog record to perform this editing process. On the basis of our present error rate of 1.05 errors per catalog record, our error-correction cost is approximately twenty-five cents per entry. This amount represents computer time only; the typist's time (about 2.3 minutes) and the proofreader's time must also be added to determine total error-correction cost.

File Formatting

Catalog records, after having been edited in accordance with the above procedure, are restructured in order to organize the material into our standard file format. In this operation a table of field locations is attached to the records for more convenient searching and information is extracted for later processing into the inverted files.

Several modifications have been made to the program that formats catalog records to adapt it to changes in the cataloging procedure and to correct faults that have been discovered during its operation. The program currently requires approximately 2.5 seconds of computer time to format one catalog record. The average catalog record requires slightly more than 500 computer words of storage in its formatted form.

Inverted-File Generation

A revised, generalized sort-merge program developed by the M.I.T. Technical Information Program (TIP) has been used for sorting the inverted files. The sorting of 3,690 subject/title terms taken from a sample collection of 326 catalog records required approximately five minutes of computer time (about 0.1 sec per item or about 1.0 sec per catalog record). The subject terms for this sample collection averaged between four and five English words; the resultant subject/title inverted file consisted of about 50,000 computer words and required 30 additional seconds of computer time to generate. The time figure for constructing the inverted file for the 650 author names (508 different names) was correspondingly smaller.

Inverted-File Searching and Listing

Searching the inverted files takes an average of about 0.2 second per search (that is, about the time for one disk access). A program to provide an online listing of selected portions of the inverted files has been written by Mr. Robert Greer and is in operation. It is currently being revised to provide a complete off-line listing, as well.

Catalog-Record Retrieving

Retrieving a particular catalog record from the catalog file requires typically from one to two seconds; however, the variation in times for our sample runs is so great that additional testing is needed.

User-System Dialog

Because the Phase-I system is used only by Intrex analysts, it was decided to keep the Phase-I user language very simple. Hence, the format is rigid and requires a professional user.

In the Phase-I dialog the system first asks for the MODE of the search (subject/title or author; exact match or prefix match). Then, in response to NAME, the user gives the subject term, title, or author name to be searched for. Finally, in response to FIELD, the user gives the field number of the field to be retrieved from the catalog file and output for each matching reference found.

After being given these three items of information the system performs the search and responds with the number of matching references found, if any, and the attributes of these references--document number, property code, and subject/title weight or author's initials. Then the given field is retrieved and output for each document. In addition, a series of time checks are reported to aid the analyst in determining how much time was taken in various stages of the search. A "quiet" mode is available when abbreviated output is desired.

Under moderate loading conditions on the CTSS system (15 to 25 users), a typical response delay time, which is the time between the end of a user's statement and the beginning of the system's response, is from two to ten seconds. This time is essentially the waiting time in the time-sharing system while other users are being serviced.

Systems Programs

Several utility procedures to perform such functions as sorting, input and output operations, and string manipulation have been implemented and added to our program library for regular use by the programmers.

THE PHASE-II SYSTEM

The broad outlines and goals of the Phase-II system were described in the preceding Semiannual Activity Report. In the following sections we shall detail some of the design features of this system.

File Organization and Generation

The catalog file will be the same as described previously for Phase I. (See preceding Semiannual Activity Report.) The inverted files will maintain the same basic "directory-section-list" organization previously described for Phase I, but the list structure will be modified to allow for phrase-to-word decompositon and stemming. A detailed description of the resulting format and the parameters involved is given in Fig. 4. It may be noted by comparing this figure with the corresponding figure (Fig. 5) in the preceding Semiannual Report that additional parameters describing (English) word endings and word positions have been added to the list format and the reference words.

Some modifications are also required in the file-generation programs to allow for exceptionally long lists and records (the "magnet" list, for example, is expected to have nearly 10,000 references) as well as to provide space within the lists and sections for updating.

Term Matching and Relevancy Measures

In the typical case, the user's query term, if more than two words long, will probably not exactly match any subject term in the catalog. This is especially true in the Intrex environment where catalog and user language is unrestricted and where catalog terms tend to be fairly long phrases. Therefore, the retrieval of pertinent documents becomes a question of making partial matches between subject term and query term and of trying to establish the degree, or relevancy, of the match for each case.

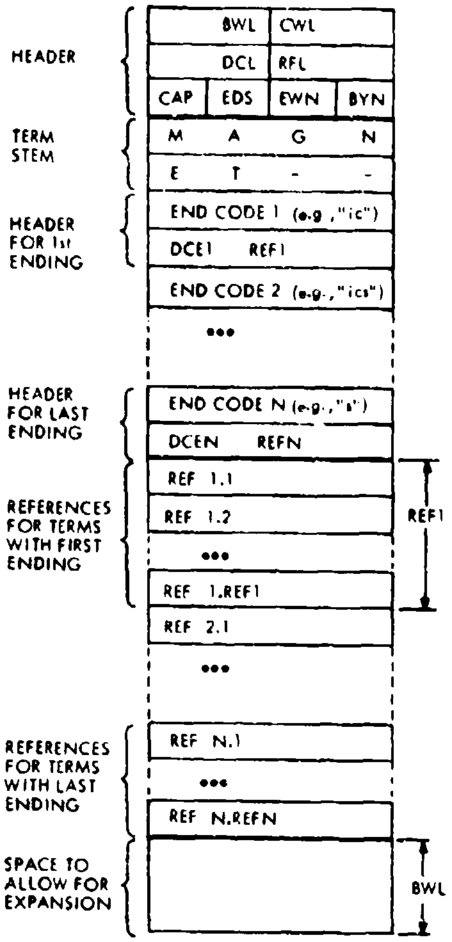
To establish a general framework in which to investigate relevancy measures, the concept of a relevancy vector has been devised. A relevancy vector is a single computer word in which we store parameters identifying the relevancy of a given document (or relevancy of a subject term in the document's catalog record) to a particular user query. There are three parts to this vector:

1. document number and, possibly, subject-term number for that document;

2. match vector--a series of bits indicating for which words in the query a match has been found in the document or subject term;

3. a cumulative relevancy measure--an integer expressing the current estimate of the relevancy between query and document or subject term.

The cumulative relevancy measure can be incremented by a variable amount for a given reference, depending on factors such as: how many words in the reference phrase match query words; whether the reference word(s) match(es) the query word(s) exactly, or only on stems; whether the reference word is in the same subject term, or merely the same document, and so forth. A list of relevancy



BWL: Number of blanks at end of list (for expansion)

CWL: Total number of computer words on list

RFL: Number of references

DCL: Number of distinct documents among references

BYN. Number of bytes in term stem (here 6)

EWN: Number of English words in term (here 1)

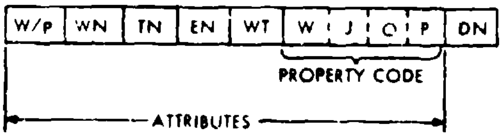
EDS: Number of different endings (here N)

CAP: Bit to indicate variable capitalization

REF1: Number of references for the work "magnetic"

DCE1: Number of distinct documents referring to "magnetic"

REFERENCE-WORD FORMAT



DN: Document number

W/P: Is term a single word (W) or full phrase (P)?

WN: Word number within phrase (for W/P = W)

TN: The term number of this term for given document.

EN: Word ending number for this reference (from 1 to N)

WT: The subject/title weight (level).

W: Is document whole work?

J: Is document journal article?

O: Does document reflect original work?

P: is document written for professional?

Fig. ~ ormat for Phase-II Subject-Term List vectors can then be developed during the searching stage and can be ordered by relevancy measure to provide an estimate of the most relevant documents. This technique also provides a means for analyzing the effects of a list of references incrementally, that is, it avoids the need to store all references before any anaylsis can be started.

User-System Dialog

The user language and aids for effective user-system interaction are being designed initally in Phase II for implementation with the IBM 2741 consoles presently in use on the CTSS system. This will allow us to initiate a Phase-II user-oriented system now while also preparing for conversion to the augmented-catalog console now under development (see Section B.3).

The user-system dialog must resolve a three-part problem:

1. Introduce new users to the system
2. Introduce past users to new features of the system
3. Remind past users how to use various features

Our aim is to develop methods that will not intimidate, or otherwise aggravate, present and potential users. It is recognized that a variety of techniques will be needed to accommodate the anticipated diversity of user preferences.

Currently three basic methods of user-system dialog are being studied. These include: a guide that would be available both online and in printed form; a branching program that directs the user, while he is working online, to system capabilities as he sees the need for them; and a fairly free-format input language which enables the user to begin negotiations with the system with a minimum of preparation on his part. Subroutines for manipulating and printing textual material, which are required for implementing all three approaches, have been written. An initial version of the guide and sample dialogs have also been prepared.

A closely related problem is that of introducing new users to the time-sharing system and to the use of existing consoles. In order to free users from the difficulties that can be associated with logging into the time-sharing system, we are planning to maintain at least one dedicated console that will remain logged-in at all times that the time-sharing system is available. We are also planning to prepare a small card, or booklet, which will introduce the potential user to features of console operation that are needed to perform such basic operations as correcting errors or interrupting system operation. This card or booklet will also show the user how to log in if he prefers to use a console that is not dedicated to Intrex.

THE PHASE-III SYSTEM

Thesaurus Generation

Mr. Florian Guertin has begun to analyze the results of applying the thesaurus-generation program devised by Richard Domercq (see preceding Semiannual Activity Report) to the Intrex data base. One fairly obvious conclusion is that the correlations possible with the long Phase-I phrases are too meager to be interesting. Therefore, thought is being given to modification of the Domercq programs to accept the Phase-II inverted files as input.

Automatic Indexing

Mr. William Kampe, in conjunction with his M.S. thesis, is continuing the development of programs to investigate methods of automatic indexing. He has written programs to determine which words in subject terms generated by the catalogers for a given document are also found in the title and/or abstract of that document. These statistics will be used to derive a dictionary of likely subject words as an aid to automatic pre-indexing of documents.

SEARCH SYSTEM IMPLEMENTATION FOR AN AUGMENTED CATALOG FOR A LARGE LIBRARY COLLECTION

We are continuing the longer range study to determine appropriate means for implementing a computer system capable of handling the search function for a collection of one million documents. Two studies are now underway. One is concerned with the physical means for storing catalog information. The other is concerned with methods for organizing large files and is independent of the implementation used.

Mass-Storage Scheme

For concreteness, we are continuing to consider a collection of one million documents, which would give rise to inverted files containing of the order of 10^9 bits and a complete catalog file with over 2×10^{10} bits. As documented in the preceding Semiannual Activity Report, our studies have shown that implementation of the inverted files can be reasonably carried out by means of an existing mass-storage device, but that the catalog file, if implemented with existing devices, would involve excessive costs. To reduce costs and to take advantage of the characteristics peculiar to a library system, we have evolved a mass-storage concept that is expected to result in significant cost reduction.

The key considerations are these:

* Access time of several seconds
* A storage cost significantly less than 0.004 cent per bit
* High bit density per unit storage area
* High bit rate in read mode

* Insignificant wear with repeated reading of same storage area
* Trouble-free operation
* Online operation in read-mode only

The basic concept we are exploring uses multitrack magnetic tape and can be described with the aid of the simplified schematic diagram of Fig. 5. The tape is

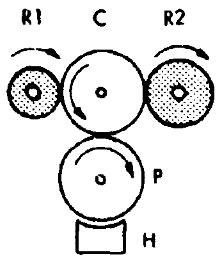


Fig. 5 Tape Drive and Read Mechanism

wound on reels R1 and R2, which have no flanges. Cylinder C is a capstan and has flanges. It is motor-driven and when turning counterclockwise as shown, tape is fed from reel R1 to reel R2, both of which are spring-loaded against C and free to rotate except for frictional drag. Observe that there is very little relative motion between the tape and the other parts that it is in contact with, so that wear is expected to be minimized. This part of the device is the Newell drive, which is finding considerable interest in the magnetic-tape field. Also shown in Fig. 5 is a printing wheel P, which, like R1 and R2, is spring-loaded against C and free to rotate. The surface of P is coated with a thin magnetic film to which is transferred the flux pattern (bit pattern) stored on the tape. The "printing" mechanism leaves the pattern stored on the tape unchanged and, because there is no relative motion between the tape and P, the tape is expected to show much less wear than in conventional in-contact reading, where troublesome wear occurs. Finally, the figure shows a head assembly H. Here the information transferr to the coating is read and then erased prior to coming in contact with another seg. ent of the tape. The head assembly may be either fixed, in which case there must be as many heads as there are tracks, or it may be movable along the direction perpendicular to the plane of the figure. Because the axis of rotation of P is fixed, provision for movable heads is not anticipated to be difficult.

The figure shows only the read operation. Writing, required in adding and modifying entries, will be performed on a separate device. By arranging for reading in both directions of motion, tape rewinding is made unnecessary.

To achieve the kind of access time which is appropriate to a catalog-file lookup, a drive would hold 200 to 400 feet of tape, which could be made two inches wide. We estimate that ten such devices can accommodate the entire catalog for a collection of the size stipulated. Because of the simplicity of each device, we anticipate that our objective of low cost will be met.

Analytical studies made so far indicate that the scheme described above is feasible. Our current work centers on planning an experimental program that will establish appropriate electrical-design parameters for the system. Of primary interest are the specifications of the thin-film coating on the printing wheel and the design of the heads which will assure that useful experimental data can be obtained.

File Organization

The manner in which files are organized affects the accuracy of retrieval, the speed of service, and system cost. In this study, we are attempting to develop quantitative relationships among these characteristics which will make it possible to find file organizations well-suited to a large catalog. It is possible to make this study device-independent, because all mass-storage devices have the common characteristic of being divisible into "quanta" which are individually addressed, and the information in each quantum is delivered to the processing equipment in its entirety. (An example of a quantum is one track of the tape in Fig. 5, or one track on a magnetic disk.)

Much of the work to date has been concerned with the isolating of parameters which determine file effectiveness: recall, relevance, access time, and file size. In order to put precise meaning into each parameter, the total document-retrieval system has been decomposed into its component parts so that individual error contributions may be studied independently. For example, to study the loss of relevance introduced by economies in file organization, it is first necessary to isolate this effect from relevance loss caused by poor translation of the user's request into index terms or caused by poor initial document-content evaluation. This definition of parameters is essentially complete now so that further work can concentrate on actual file organization.

The particular aspect of file organization being studied involves the ordering of documents into groups with similar content (clustering), in order to facilitate file access. The objective is not only to study practical techniques for doing this, but also to estimate how successful such grouping is, as measured by file cost-performance figures. Emphasis is on determining the effects of document redundancy, namely the repeated appearance of documents in the file to improve access time and relevance. Computer simulation will be used to estimate file cost as a function of library size, organization technique, and characteristics of the document collection.

3. THE DISPLAY-CONSOLE SYSTEM

Staff Members

Dr. D. R. Haring
Mr. J. E. Kehr
Professor J. K. Roberge

Graduate Student

Mr. P. F. McKenzie

SUMMARY

In the preceding Semiannual Activity Report a complete system design for the augmented-catalog display console was described. During the present reporting period, construction of this system was begun. A system test set has been designed, constructed, and made operational. This test set simulates the digital signals that occur in the display system and is an invaluable tool for testing the performance of the individual digital elements contained in the system and the system itself. The construction of the logic and the electronic circuits for the user console is nearly completed and much of the mechanical work is well underway. Minor changes have been made in the system configuration described in the preceding Report in order to improve system performance and to take advantage of special features of the detailed properties of the logic building blocks being used.

Work continues on the software required by the local processor, some of which will be described here. Experimental investigations that are currently being conducted on the display unit are also presented.

SYSTEMS CONSIDERATIONS

Figure 6 shows a complete functional block diagram of the user console described in the preceding Activity Report. Details of its operation, not previously reported, are presented in the paragraphs that follow. The specific realizations of the system digital elements that are employed in the user console were chosen in lieu of other possible realizations because those chosen are the most economical when presently available integrated circuits are used. Although these realizations may very well change with the continuing developments of digital integrated circuits, the user-console system configuration itself should change very little.

The connection between the user console and the buffer/controller (B/C) is shown in the lower right part of the figure. The two interconnecting coaxial cables between the console and the B/C that were shown in the preceding Activity Report have been replaced by a single interconnecting coaxial cable and directional couplers on either end of the cable. The directional couplers are inexpensive compared to the required length of coaxial cable. Digital signals on the cable in both directions are bipolar to enhance noise immunity and simplify the clock regenerator.

The output from the directional coupler in the console drives an integrated-circuit comparator which in turn drives the clock regenerator and the input shift register. This data path carries the data from the drum memory in the B/C to the viewing cathode-ray tube (CRT) in the console. An additional signal from the processor in the B/C to the console--called "Interrogate"--is a single burst of sine waves that tells the console the processor is ready to receive the data stored in the data shift register and hence initiates the data transfer from console to B/C.

The output from the clock regenerator is used to synchronize all operations in the console with those in the B/C. It is used to synchronize a sine-wave vertical-deflection signal for both the character generator and the viewing CRT in order to prevent flicker on the viewing screen.

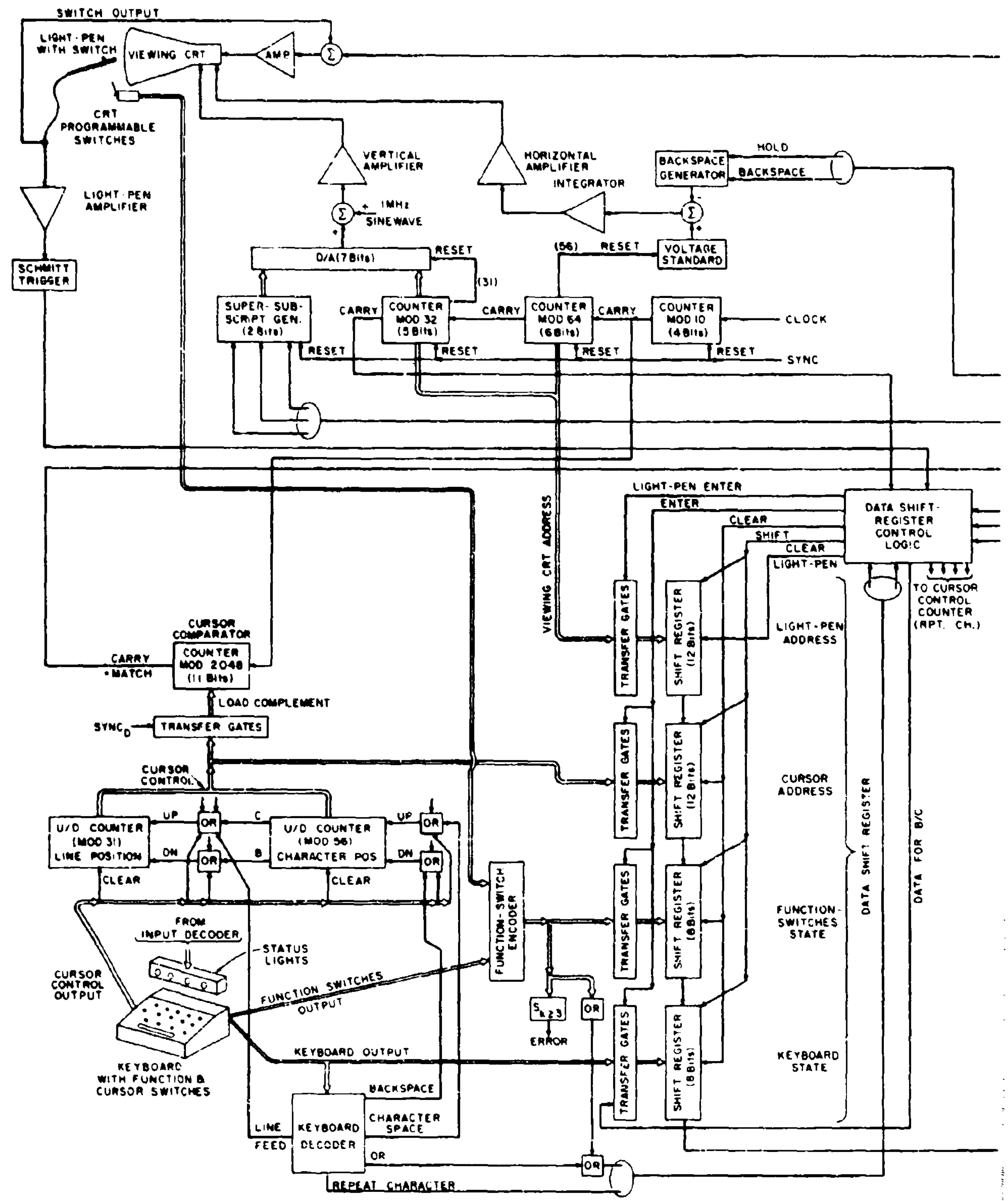
When the console is first turned on, or if for some reason synchronization between the console and the B/C is lost, the console is able to lock onto synchronization signals from the B/C in the following manner: The rate and the phase of the individual bits contained in the input data received by the console are determined by the clock regenerator. To determine the 10-bit code groups corresponding to the characters sent by the B/C, the first bit of each of these code groups must be identified by the console. This identification is performed by the "word" synchronizing technique described in the next paragraph.

Inspection of Fig. 6 reveals that when the last stage of the input shift register (stage number 10) is set to the 0 state the shift register is immediately set to the state 11...1. Hence, if the console is out of word synchronization, the first 0 in the input string of bits from the B/C will be defined as the first bit (bit 1) of a code group. In general, this 0 will not be bit 1 of a code group; but if two identical code groups of 01...1 are received by the console in adjacent 10-bit word positions, then one can guarantee that by the time the second 01...1 code group is received, the console will be in word synchronization. Such a pair of code groups is recorded on the drum track driving the console during the 32nd line of textual material, which is the invisible line on the console devoted to the vertical retrace of the viewing CRT.

One more synchronization technique is required, namely that of "frame" synchronization. This synchronization insures that character position j on the viewing CRT corresponds to character position j on the drum memory. Frame synchronization is effected by sending an ASCII sync-code group at the end of the 32nd line. The code group is detected by the input decoder, which then sends signals to the appropriate logic elements in the console.

Console Operation

Let us now describe the operation of a completely synchronized user console. Referring again to Fig. 6, observe that the input shift register drives both the input decoder and the character-generator buffer register. The former detects the presence of control characters such as sync, character-set change, backspace, superscript, subscript, and status-light excitation. The character-set-change signal is required because the console will operate on an expanded set of characters


SWITCH OUTPUT
LIGHT-PEN WITH SWITCH
VIEWING CRT
AMP
CRT PROGRAMMABLE SWITCHES
VERTICAL AMPLIFIER
HORIZONTAL AMPLIFIER
INTEGRATOR
BACKSPACE GENERATOR
HOLD
BACKSPACE
LIGHT-PEN AMPLIFIER
1MHz SINEWAVE
SCHMITT TRIGGER
D/A (7 Bits)
RESET
(31)
(56)
VOLTAGE STANDARD
SUPER-SUB-SCRIPT GEN. (2 Bits)
CARRY
COUNTER MOD 32 (5 Bits)
COUNTER MOD 64 (6 Bits)
COUNTER MOD 10 (4 Bits)
CLOCK
SYNC
LIGHT-PEN ENTER
ENTER
CLEAR
SHIFT
LIGHT-PEN
DATA SHIFT-REGISTER CONTROL LOGIC
TO CURSOR CONTROL COUNTER (RPT. CH.)
VIEWING CRT ADDRESS
TRANSFER GATES
SHIFT REGISTER (12 Bits)
LIGHT-PEN ADDRESS
CURSOR COMPARATOR
COUNTER MOD 2048 (11 Bits)
CARRY •MATCH
LOAD COMPLEMENT
SYNC$_D$
TRANSFER GATES
CURSOR CONTROL
CURSOR ADDRESS
DATA SHIFT REGISTER
DATA FOR B/C
U/D COUNTER (MOD 31) LINE POSITION
UP
DN
OR
C
B
U/D COUNTER (MOD 56) CHARACTER POS
CLEAR
FUNCTION-SWITCH ENCODER
SHIFT REGISTER (8 Bits)
FUNCTION-SWITCHES STATE
FROM INPUT DECODER
STATUS LIGHTS
CURSOR CONTROL OUTPUT
FUNCTION SWITCHES OUTPUT
$S_{\geq 23}$
ERROR
KEYBOARD WITH FUNCTION & CURSOR SWITCHES
KEYBOARD OUTPUT
KEYBOARD STATE
BACKSPACE
CHARACTER SPACE
LINE FEED
KEYBOARD DECODER
OR
REPEAT CHARACTER

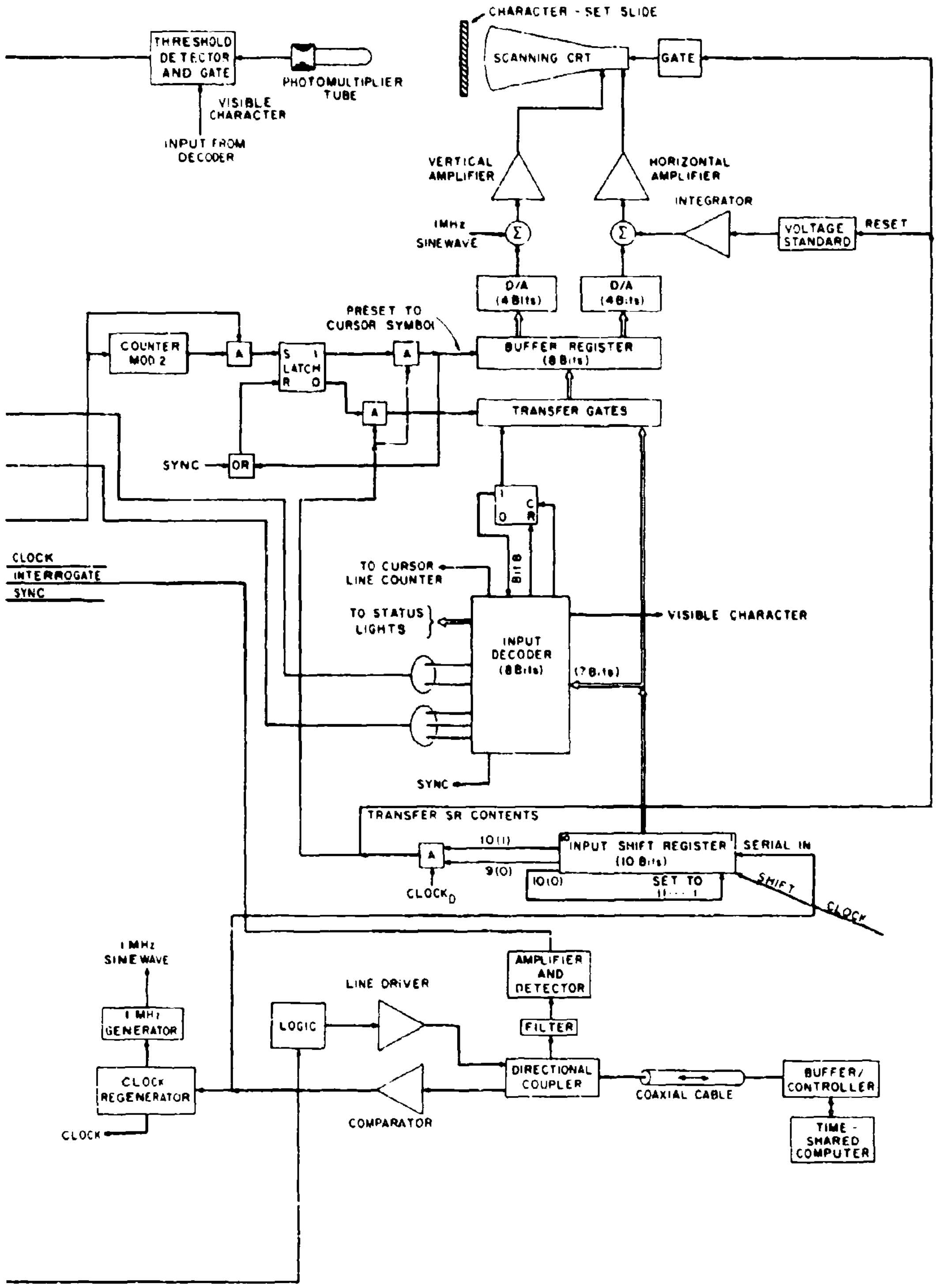


Fig. 6 Functional Block Diagram of the Augmented-Catalog Console

such that the seven bits in the 10-bit code group used to define the character do not provide enough code groups for all desired characters. The console operates initially in the standard ASCII set of 128 characters. If a character from the second set of 128 characters is desired, the 7-bit code defining that character must be preceded by the code defining the shift to the new character set. The occurrence of this shift character sets the complementing flip-flop (FF) shown above the input decoder to the 1 state. The state of this flip-flop is the 8th bit combined with 7 bits from the input shift register that drives the character generator. To return to the standard ASCII set, one must repeat the shift character, which occurrence sets this complementing FF to the 0 state.

The contents of the character-generator buffer register are changed when stages 9 and 10 of the input shift register are in the states 0 and 1, respectively. At this time the entire code group is in the input shift register (except for the final 1 which need not be decoded) and is ready to be decoded.

The binary word in the character-generator buffer register is converted into two analog signals that are used to find the position in the character-set slide at which the character whose code this binary word represents is located. Once located at the correct character position, the character is scanned by means of a linear ramp on the horizontal axis and a sine wave on the vertical axis. These signals are phase-locked with the corresponding signals on the viewing CRT so that if the character being scanned is a visible one, the output from the photomultiplier tube (PMT) drives the intensity control of the viewing CRT to reproduce the desired character.

When the character address of the viewing CRT is equal to the address of the cursor, then the character generator alternately displays the symbol for the cursor and the symbol recorded on the drum at this address. This alternation is produced by the Modulo-2 counter and latch, (located to the left of the character buffer register). The counter and latch either directly set the character buffer register to the binary code for the cursor or transfer the contents of the input shift register into the character buffer register.

In the upper left quarter of Fig. 6 is found a series of three counters. The Modulo-10 counter is driven directly from the clock and thus puts out a carry pulse for every 10-bit code group defining a single character. The Mod-10 counter drives a Mod-64 counter, the content of which specifies the character position of the line that is presently being displayed on the viewing CRT. During states 56 through 63 of this counter, the integrator that generates the linear horizontal sweep signal is held at a voltage corresponding to the left end of lines displayed on the CRT. This period of eight character positions provides sufficient time (approximately 80 microseconds) for the deflection circuits of the CRT to retrace from the end of the present line to the beginning of the next line. Note that the integrator is also driven from the back-space generator. This generator, on command from the input decoder, produces a pulse on the integrator input such that the viewing CRT beam remains fixed in position if a control character, such as superscript, is decoded (this is called "hold"), or the generator produces a pulse on the integrator input such that the viewing CRT beam actually goes back one character position if a control character, such as backspace, is decoded. One character time interval (approximately 10 microseconds) is allowed for completion of either of these operations.

A Mod-32 counter to specify the line that is presently being displayed on the viewing CRT is driven by the carry output of the Mod-64 counter. A digital-to-analog (D/A) converter on the output of this counter is used to generate the vertical deflection of the viewing CRT. When the Mod-32 counter reaches state 31 the D/A converter is reset to zero voltage in order to allow the deflection circuits of the viewing CRT the time of a full line (approximately 640 microseconds) to do a vertical retrace from the bottom to the top of the screen, and a horizontal retrace from the right to the left of the screen.

In addition to the D/A converter from the Mod-32 counter, the vertical-deflection circuit is driven from the Super-Sub-Script generator. The latter generator drives D/A converters from the outputs of a pair of Mod-4 counters to generate vertical deflections of plus and minus 1/4th to 3/4ths of a line width. This makes it possible to present super-super-super-scripts, sub-sub-sub-scripts, and combinations thereof.

All the above counters are reset to 00...0 on the occurrence of the sync signal during line number 32 at the end of the frame in order to make the deflection circuits of the viewing CRT automatically frame synchronous.

The data shift register in Fig. 6 has 40 bits that are partitioned as shown. The operation of this shift register is under control of the data-shift-register control logic, which is a sequential circuit and operates roughly as follows. Assume that the console has just been turned on and no manual inputs have been created. Under these conditions the control logic is in a stable state and remains there until one of the following three things occur: (1) A function switch (with fixed or programmable labels) is actuated; (2) A keyboard switch is actuated; (3) The light-pen switch is depressed and the light pen senses light from the viewing CRT screen. If either (1) or (2) occurs, the control logic sends an ENTER signal to the data shift register, which transfers the contents of the keyboard, the function-switch encoder, and the cursor control to the respective shift-register bits, sends a 1 signal to the processor in the B/C to indicate the console has data for the processor, and causes the control logic to enter a new state such that no new manual inputs to the console will influence the content of the data shift register. This stable state continues until the processor sends an INTERROGATE signal to the console. Upon receipt of this signal the control logic goes to a new state which causes the content of the data shift register to be transferred to the B/C. At completion of the transfer, the data shift register is cleared and the console is ready to repeat the process when a new manual input to the console is created. If the operator of the console continues to actuate the switch that created this sequence of operations, one of two things occurs. If the switch creates one of the "repeat" characters (a character that can be repeated by simply holding down the switch generating that character), then, after a period of time determined by the control logic, the sequence of events described above is repeated and this sequence continues to be repeated as long as the switch is actuated. On the other hand, if the actuated switch does not create one of the repeat characters, then the control logic remains in a stable state in which no data can be sent to the B/C until that switch is released, and a switch is actuated after the release.

The data-shift-register control logic provides special signals for the light-pen logic. The content of the Mod-32 and the Mod-64 counters driving the deflection circuits of the viewing CRT are used by the light-pen logic to indicate the instantaneous address of the character position being identified by the light pen. Regardless of the state of the other console switches, the content of these counters is transferred to the appropriate part of the data shift register by the control logic when light is sensed by the light pen and the light-pen switch is pushed. This transfer occurs unless the control logic is processing a previous switch actuation, in which case no transfer takes place. To maintain the most recent light-pen position in the data shift register, when the control logic is not processing a previous switch actuation, the shift-register bits corresponding to the light-pen address are cleared by the sync signal.

The function-switch encoder in the lower left quarter of Fig. 6 encodes the function switches into a 2-out-of-8 code, that is, a code in which every function switch has an 8-bit code in which precisely 2 of the 8 bits are 1. Such a code is used to simplify the encoder logic and the decoding performed in the processor, and it allows a maximum of 32 function switches, which should be an ample number.

The cursor address is determined by the content of a Mod-56 up/down binary counter to specify the character position in the line, which is specified by the content of a Mod-31 up/down binary counter. These two counters are also shown in the lower left quarter of Fig. 6. The address of the cursor can be changed in a variety of ways. First, there are four control switches on the keyboard which can move the cursor up, down, left or right. Second, there are the standard teletype-like commands of back-space and line-feed. Third, whenever a character is typed on the keyboard, the cursor is automatically moved one space to the right. Finally, a fourth way the cursor can move is under command of the processor when the console is in the dialog mode.

In order to reduce requests to the processor in the B/C, many changes in the cursor address are not sent to the B/C, because the address is transferred to the appropriate bits of the data shift register only when needed. The need occurs typically whenever a function switch or the keyboard is actuated.

To display the position of the cursor, when the address of the cursor corresponds to the character address that is being displayed on the viewing CRT, the cursor symbol and character are displayed alternately, as described previously in this section. The correspondence of the two addresses is determined by loading the TWO's complement of the cursor address into a Mod-2048 counter at the instant the first character address of the viewing CRT occurs. This counter is advanced one count for every character position. Thus, since an n-bit binary number plus its TWO's complement is equal to 2^n, a carry from the Mod-2048 counter signifies that the two addresses are equal.

DISPLAY-SYSTEM SOFTWARE

The Varian Data Machines 6201 processor in the buffer/controller serves as a system monitor designed to control data flow among the elements of the Intrex system configuration. Essentially it must service the multiple Intrex consoles and link them to the central time-shared computer system which is remotely located. It also provides user communication with the hard-copy acquisition hardware. For purposes of this section, the simplified system configuration shown in Fig. 7 will be useful.

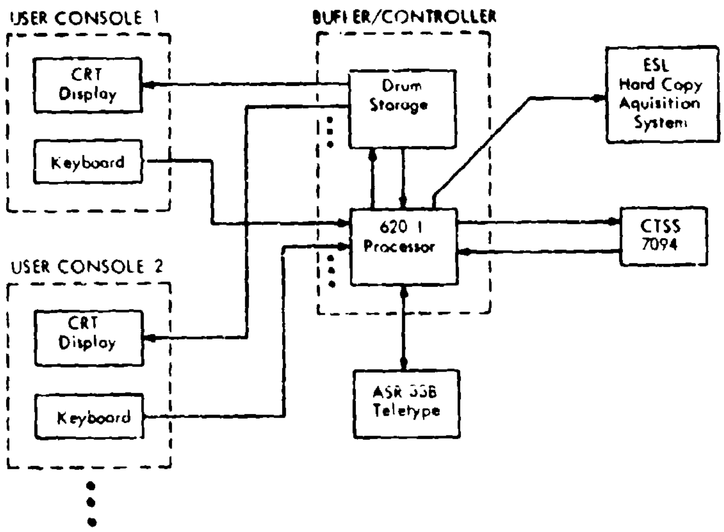


Fig. 7 The Display-Console System Configuration

The 6201 processor requires a flexible monitor program to control data flow efficiently and to provide linkage to subprograms that provide special services for the user console. User actions and current system conditions are sensed by the monitor which calls into action the appropriate subprogram to initiate a particular machine response.

The development of the 6201 software is proceeding concurrently with hardware development of the controller and the Intrex user console. As these efforts proceed, hardware-software tradeoffs are constantly evaluated and design and programming features are modified with the usual convenience, flexibility, cost, simplicity and speed considerations in mind.

To check out software for an evolving system presents some interesting problems. Each hardware element of the system configuration is an integral part of the monitor software. Yet some portions of the final configuration will be available before others. Because of the time involved in software preparation and check-out, it is desirable to start exercising programs as soon as possible. Two approaches are apparent. One is to break the software into parts which can be exercised with a partial system configuration. This has been done to a limited degree in a set of utility programs that can be used to exercise and test both hardware and software piecemeal. These utility programs will serve as a backbone for system debugging and diagnostic programs to be used by the 6201 system programmer in later modifications to the Intrex operating experiment.

This utility program approach does not allow an adequate check-out of the interaction of system elements. The second approach is more interesting and involves a form of simulation. The basic 6201 processor will be available first and will be exercised to check the machine instruction set and control features. The manufacturer-furnished assembly program and debugging package will be exercised, and modifications and procedures will be developed to accommodate special Intrex system features. At this time, it will be possible to assemble programs which cannot be run because of missing system elements; hence, device simulation seems a logical next step. The 6201 will be programmed to simulate its missing element in a limited but useful way.

In light of the Intrex configuration, asynchronous operation of the system elements is of vital importance. The operation of the monitor results in a highly interconnected multiple-queue process. As a result, any meaningful program check-out must take into account these asynchronous random processes. The simulated device action times will differ from the actual real-device action times, so a simulated time must be kept as it is in standard simulation programs.

A nice handling of this simulation would be an interpretive process in which the 6201 program would be assembled in the exact form in which it would be assembled for the complete system. The interpretive program would call up each program instruction, decode it and, if the instruction was one that was allowed by the existing configuration, it would be executed. If the interpreter found an instruction that was not allowable because the device called was not in the system configuration, then an appropriate subroutine would be called and simulation would occur. This solution should result in a debugged and in some degree a verified assembled program that could be used directly in the final system configuration (with modifications for evolving design changes). The drawback lies in the programming effort in writing this flexible interpretive-simulator. The running time would be long but that is no real problem since the 6201 is used exclusively for Intrex, and simulation is for program check-out, not for use under operating conditions.

An alternate approach has been taken. Some of the functions performed automatically by the interpretive program will be performed by the programmer as he checks his programs. He will insert special calling sequences to isolate the simulated operations; he will do the decoding to determine what must be simulated and he will handle some of the simulated-time updating chores.

This technique will result in a checked-out program that will have to be modified (patched or reassembled) to remove the inserted calling sequences which are foreign bodies in the operating program. This compromise solution is being followed currently. If the more general solution appears practical, the compromise program can be adapted to that goal.

The ability to simulate missing system elements is appealing not only for program check-out but also for debugging and diagnostic efforts when the system is in full operation. It will make possible a reversion to some minimal subsystem when necessary to isolate system faults.

A first version has been written for the simulation of the central time-sharing system query-response mode and for the character mode of drum read and write.

A program has been written to permit use of the ASR-33 teletype input device of the 6201 in the same manner as the IBM 2741 is currently being used by Intrex to communicate with the central time-shared processor. In this mode of operation the 6201 becomes invisible by virtue of its program. A version of this invisible-mode program will be used to simulate the Intrex console input.

CHECK-OUT SIMULATION OF ASYNCHRONOUS DEVICES

The technique for modifying an existing program for check-out through employment of the asnychronous device simulator is presented schematically in Fig. 8.

The correct use of the asynchronous devices requires that they be tested for their availability prior to their engagement in operations. If the programmer fails to do this, he will lose data through engagement of devices that are not free. The simulator will behave in an analogous manner.

In the case of incorrect coding for asynchronous operation, the incorrect code may not yield results on the real device that are identical to results obtained on the simulated device. Even so, incorrect coding should be detectable by careful program checking and testing. The level of complexity of simulation to obtain identical results in both cases would be too great for present goals.

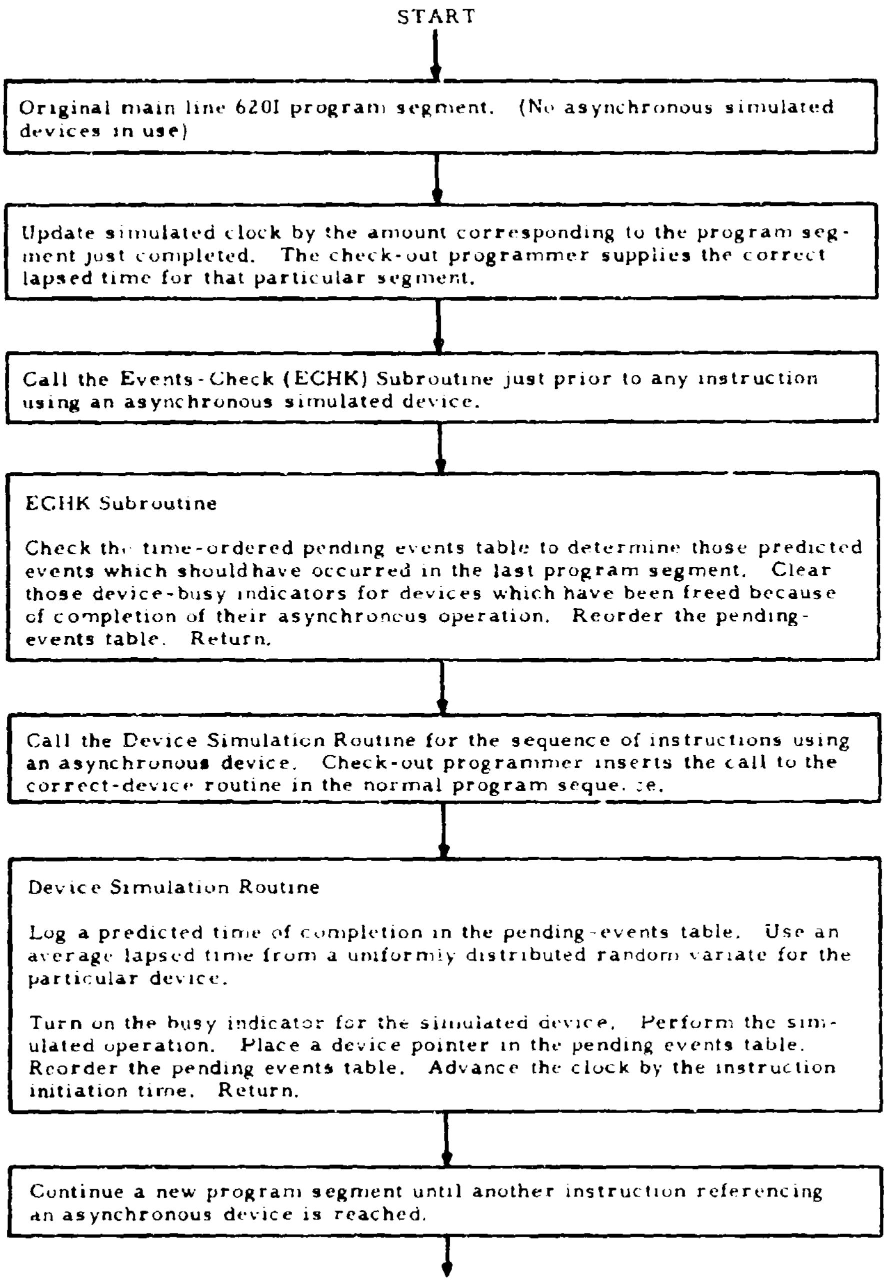


Fig. 8 Statement of Program Modification to Effect Asynchronous Device Simulation

In the programmed use of data delivered directly to core by an asynchronous device, it is also necessary to check for completion of delivery prior to use. Incorrect coding will not yield identical results on the simulated and the actual devices; in fact, repeating the same sequence of incorrect coding on the actual device will not necessarily yield identical successive results due to the asynchronous nature of the device.

THE CHARACTER-GENERATOR AND DISPLAY-TUBE CIRCUITRY

A major effort has been placed on further development of the flying-spot-scanner character generator discussed in earlier progress reports. Figure 9 shows

Fig. 9 Example of Output from Character Generator

an example of characters generated with the present equipment and displayed on an entertainment-quality cathode-ray tube. The writing rate and character size achieved to date are sufficient for 1000-character displays. Further refinements of the technique should permit achievement of the design goal of approximately 1800 characters per display.

Investigations are currently underway to increase the speed with which the display tube is blanked and unblanked, since this speed is the present limitation on the number of characters that can be displayed. Several approaches to this problem are under consideration. It seems necessary to use a P-16 phosphor for the character-generator CRT to achieve the required generation rates. Disadvantages of this phosphor include a relatively high susceptibility to burning and the need to design any optical components used in the system for ultraviolet transmission. These disadvantages do not seriously limit our system, however, since the phosphor can be protected by appropriately blanking the CRT beam and since lenses have not proved necessary for the character generator.

The effect of character style on speed of response is also under investigation. Because of phosphor decay time, the width of the lines that make up the characters on the character-set film affect character generation rate. Investigations show that the speed of response can be improved by appropriately selecting these line widths. Current investigations seek to determine the optimum line widths.

A high-voltage supply for the display and character generator cathode-ray tubes has also been developed. Since deflection sensitivity is related to the accelerating potential applied to these tubes and since short-term changes in deflection sensitivity result in blurred character images, the high-voltage supply must be well regulated to preserve character fidelity. A unique design which uses an automobile-type ignition coil as a flyback transformer and which provides a 10-kv output with short-term regulation within 0.1 percent of rated voltage has evolved.

DISPLAY-CONSOLE CONSTRUCTION

The construction of the electronic and mechanical equipment for the first Augmented-Catalog Console was initiated during the reporting interval. Most of the display electronics used in the first console will differ in only minor respects from prototype circuitry developed to establish feasibility. The modifications which have been introduced during the present redesign will reduce cost and/or improve ease of assembly. Advantage is being taken of advancements in componentry which have been made during the past year. Approximately 50 percent of the display electronics and approximately 75 percent of the digital logic has been constructed in final form.

An intensive design effort is being applied to the human-engineering aspects of the display console, since it is imperative that a user's initial contact with the console, the only part of the Intrex Catalog System which the average user sees, be a pleasant one. The objective is to retain sufficient flexibility in the initial console to permit effective user evaluation of various options, while simultaneously maintaining a finished look for the console.

The display console will take the form of a two-pedestal desk. The right-hand pedestal will house electronics; the left-hand unit will be free for storage of a user's personal items. The display tube is located directly in front of the user; it can be moved vertically and horizontally, and it can be tilted to accommodate user preference. The CRT programmable switches are located at the bottom of the display tube. The keyboard will be connected to the console by means of a flexible cable so that it may be positioned anywhere on the surface of the desk.

The movable display-tube mount has been completed. This unit is counter-weighted so that its position can be altered with minimum effort. The display is normally locked firmly to its support member. The locking mechanism is released for display adjustment by means of a single pushbutton operated by the user.

C. THE TEXT-ACCESS PROGRAM

Staff Members	Graduate Student
Dr. U. F. Gronemann	Mr. A. Jagodnik
Mr. D. R. Knudson	
Mr. P. R. Scott	
Mr. S. N. Teicher	

1. SUMMARY

During the past six months the Text-Access group was engaged in two major activities: The first of these was the continuation of the measurements of modulation transfer functions of the experimental microfilm-facsimile system, coupled with further improvements and refinements to the system. The second was the design of the first experimental text-access system.

Excellent reproductions of microfilm text images are now being obtained in the microfilm facsimile system. Modulation-transfer-function curves of individual components of the system, and of the overall system, have been obtained at various parameter settings. A curve for the complete system under typical operating conditions shows an on-axis limiting resolution of 1200 cycles per frame height. Subjective evaluations of the effect of the MTF, as well as that of the number of scan lines, have been carried out, using special test slides containing samples of text and characters. The effect of slow phosphor decay in the scanner has been noted but not yet corrected.

The first experimental remote text-access system is projected for completion during the summer, 1968. It will operate in conjunction with the augmented catalog and provide access to the full text of the same 10,000 documents that constitute the catalog. Text will be stored on microfiche cards in an automatic retrieval and transmitting station. Access will be through three receiving terminals, one producing film copy and two containing stored displays. Provision for paper copy will also be made.

2. EXPERIMENTAL MICROFILM FACSIMILE SYSTEM

The modulation-transfer-function (MTF) measurements of the components in the experimental microfilm facsimile system, described in the preceding Activity Report, have been utilized to derive the system MTF under various operating conditions.

The configuration for the flying-spot scanner consists of a 3.5-in. by 2.5-in. cathode-ray-tube (CRT) raster projected by the lens onto the microfiche at a reduction ratio of 5.4 to 1 and an f-number of 5.6. The receiver CRT raster is identical to the scanner raster and is projected onto 35-mm film at a 3-to-1 reduction ratio by a lens set at f 2.8. The component MTF curves for the system under these conditions at a point on the optical axis and in the vertical-scan direction are shown in Fig. 10. Spatial frequency is plotted in cycles/page because these units are invariant with the size of the image at various stages in the system. Cycles/page may be converted to cycles/unit length for any component by dividing by the vertical size of the page image at the location of that component. The curves illustrate the relative contribution of each component to the overall-system response.

From Fig. 10 it is obvious that the scanner lens is the leading contributor to the degradation of image resolution. One reason for this degradation is that the page image is smallest in the image plane of the scanner and hence, the highest resolving power in cycles/unit length is required for this lens to achieve a response comparable to the other components. A second reason is that the near-ultraviolet spectrum of the P-16 scanner CRT reduces the lens performance, as described in the preceding Activity Report. Our lens is corrected for the spectrum of P-11, but P-16 phosphor is required for the flying-spot scanner because of its relatively rapid decay.

In general, a lens MTF is a function of both position with respect to the lens optical axis and direction in which the spatial frequency is measured. Also, imperfect dynamic focusing causes a variation in the CRT-spot size over the raster, and the effect of the video channel differs between the horizontal and vertical scan directions. For all these reasons the system MTF is also a function of position with respect to the optical axis and direction in which spatial frequency is measured. For our system, measurements were made at five positions, corresponding to the center and the four corners of the CRT raster, and in two directions, corresponding to the horizontal-and vertical-scan directions. Within the measurement accuracy, the results show that the lens MTF is the same at each of the four corners, but the difference between the on-axis MTF at the center of the CRT raster and the off-axis MTF at the corners is significant. Also, the difference between the lens MTF in the horizontal- and vertical-scan directions at each position is negligible.

The MTF for the overall system in the vertical-scan direction is shown in Fig. 11 for the on-axis position. A relative response of three percent is generally considered to define the limiting resolution point. In Fig. 11 this limit implies that approximately 1200 cycles/page on-axis can be resolved by the system. The horizontal-scan lines of the CRT raster provide, in effect, spatial samples of the scanned microfiche image along the vertical-scan direction. The sampling theorem from linear-systems theory can be used to relate the required number of scan lines to the spatial-frequency spectrum of the transmitted image. If it is assumed that the system MTF is sufficiently broad to pass the spatial-frequency spectrum required for reproducing a class of images with acceptable quality, a sampling rate of twice the maximum frequency in that spectrum allows faithful reproduction of the spectrum from the sampled points. If the number of scan lines violates the sampling theorem,

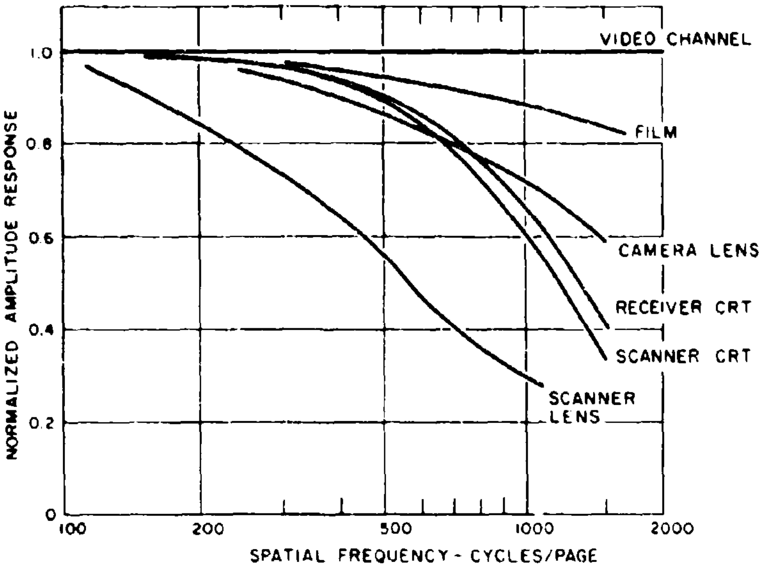


Fig. 10 The MTF's of the Various Components of the Experimental Microfilm Facsimile System

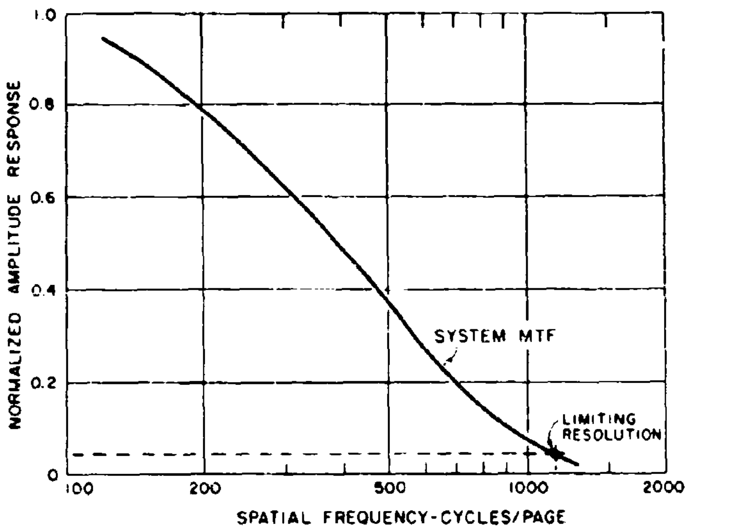


Fig. 11 The MTF of the Overall Experimental Microfilm-Facsimile System

spurious components will exist in the reproduced image and will affect the image quality. In practice, the required number of scan lines can best be determined empirically by subjective evaluations of selected test images transmitted through the system over a range of scanning lines.

The system MTF in the horizontal-scan direction is affected by two additional factors, the frequency response of the video channel and the phosphor-decay rate of the scanner CRT. The spatial variation of the image intensity of the receiver CRT is a result of the time variation of the video signal, and thus the video-frequency response limits the spatial-frequency response of the system. An equivalent MTF for the video channel is derived by dividing the time frequency of its response function in Hertz by the appropriate scan rate in units of length/sec. Because of the large difference between the horizontal- and vertical-scan rates, the video-frequency response is significant only in the horizontal-scan direction. At the scan rates required for the transmission of a single frame of 2000 scan lines in one-half second, the 3-db point of a 4.5-MHz bandwidth channel occurs at a spatial frequency of approximately 1125 cycles/page in the horizontal-scan direction, whereas the same point corresponds to 2.25×10^6 cycles/page in the vertical-scan direction. Thus, a 4.5-MHz channel contributes some loss in system response along the horizontal-scan direction, but its equivalent MTF in the vertical scan direction is essentially unity for the range of spatial frequencies under consideration.

At high scan rates, phosphor persistence causes a tail on the trailing edge of the CRT spot. In the receiver the phosphor persistence does not affect system resolution. However, the flying-spot scanner depends upon the CRT spot being nearly a point light source and its resolution in the horizontal-scan direction can be degraded significantly by the finite decay time of the scanner-CRT phosphor. At the scan rates required for one-half-second transmission per page, the half-amplitude width of the scanner-CRT spot is more than 50 percent greater in the horizontal direction than in the vertical direction. Preliminary experiments with a nonlinear circuit in the video channel indicate that compensation for this effect is possible, but the signal-to-noise ratio at the output of the flying-spot scanner may be insufficient because of the sensitivity of the compensation circuit to noise. The effects of phosphor-decay time can be avoided by slowing the transmission time per page to approximately two seconds, but further investigation of compensation techniqes is required to achieve a one-half-second transmission time per page.

A primary objective of the experimental system is to establish specifications for the system components required for a remote text-access system. Experiments have been conducted in which selected text was transmitted with different system MTF's and different numbers of scan lines. The system MTF was varied by changing the f-number settings of the scanner and receiver lenses. The horizontal-scan rate was reduced to 1250 in./sec for these tests to avoid the effects of phosphor decay.

An attempt was made to correlate subjective evaluations of the image quality with the system MTF in order to establish a quantitative measure of the system resolution requirements. An IBM 9922 document viewer with an enlargement factor of 16 was used to view the 35-mm film images of selected test targets transmitted through the experimental system. The maximum limiting resolution of the system, including the viewer, is approximately 1000 cycles/page along the vertical-scan directions. This resolution with 2000 scan lines provided good-quality images for a typical page from a technical journal microfilmed at an 18-to-1 reduction ratio. A two-column page with approximately 70 lines per column and an average of 8 to 10 words per line was chosen as a typical page.

Evaluations of transmitted images of text of various type sizes indicate that lower-case letters at least as small as 0.05 mm can be transmitted through the system. Microfiche images with letters of this size were scanned and the transmitted images were legible when viewed on the IBM viewer. Further evaluations under more controlled conditions are required to establish the limits of legibility as a function of the system MTF.

Our tests have demonstrated the difficulty in determining definitive requirements for the resolution of an image-transmission system. Subjective measures of image quality are influenced by many factors in addition to resolution. These factors include contrast, information content of the image, graininess and noise content of the image, and so forth. However, at least 2000 scan lines, with comparable system resolution, appear to be needed for remote reproduction of technical documents of average quality and containing the subscripts, superscripts, and mathematical symbols that frequently appear in these texts. At 1500 scan lines, the resolution of our receiver CRT and camera is sufficient to allow individual scan lines on the 35-mm film image to be resolved. The presence of the scan lines gives an objectionable appearance to the image. These lines could be merged by decreasing the receiver resolution, but at the expense of degrading image quality. It is possible that a higher system resolution may be required if extremely small characters or poor-quality images are frequently encountered in microfiche inputs. Further evaluations of acceptable image quality will be derived through analysis of user reactions to our experimental text-access system.

Our MTF studies have been completed and our results are being prepared for publication.

3. DESIGN OF THE TEXT-ACCESS SYSTEM

SYSTEM CONSIDERATIONS

The experimental text-access system should become operational during the summer, 1968. Its salient features will be an automatic microfiche storage-and-retrieval device capable of accommodating 750 microfiche and of being operated under computer control, a flying-spot scanner for converting microfiche images to video signals, a wideband transmission system, two types of receiver stations and a spare station, and the necessary control logic to access documents through the augmented-catalog console. One receiver station will provide microfilm as its output, a second station will produce a visual display of text on a storage tube, and a third station will be available for installation of other forms of output equipment as such equipment becomes available.

The document collection for the experimental system will be the full text of the documents in the augmented catalog. We shall, therefore, have available both catalog and full-text information at stations which are remote to the central time-sharing computer.

The overall configuration of the system is shown in Fig. 12. The system consists of: the central station which includes the automatic microfiche storage-and-

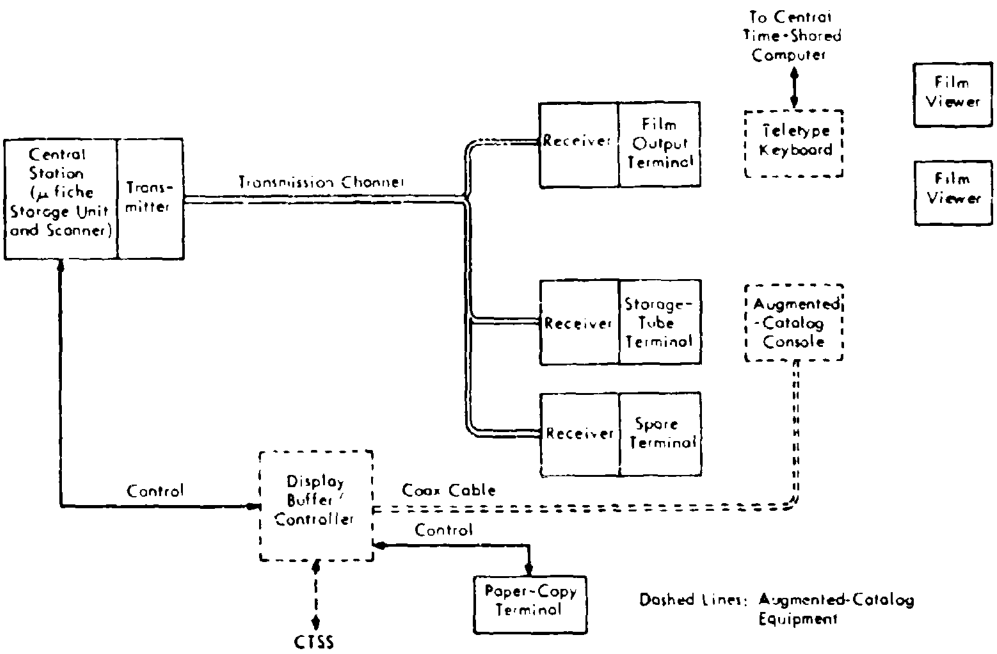


Fig. 12 Simplified Block Diagram of Experimental Text-Access System

retrieval unit, a flying-spot scanner, and driving, amplifying and control circuitry; the transmission network which comprises a transmitter at the central station, a wideband transmission link, and a receiver at each terminal; and the group of three output terminals consisting of an automatic film-output station, a Tektronix storage-tube display, and a third station for future experimental purposes. A paper-copy facility is also being planned, but its form has not, as yet, been determined.

The text-access system will be controlled by the augmented-catalog buffer/controller unit (Section B.3) through its connection to the central-station control unit. Control of the text-access system will reside mainly in the buffer/controller and appropriate software to establish this control is planned. Thus, for example, if a user requests the transmission of an entire document, the buffer/controller will remember the number of pages to be transmitted and keep track of the access and transmission operations so as to issue an appropriately timed command for each page. The access number of a document will occupy a field in its augmented-catalog entry which is stored in the central time-shared computer. Document number will be retrieved automatically whenever a document title is retrieved and thus will be available to the buffer/controller for issuance of the text-access command. Our decision to connect the text-access system to the augmented catalog buffer/controller rather than to the central computer is based on our belief that the full-text accessing process could be slowed down by the wait-time of our present time-sharing computer.

Requests for full text of documents will be entered by users on the same keyboards as those used for retrieval of catalog information, that is, either the augmented-catalog-console keyboard or the teletypewriters used in conjunction with the central time-shared computer. In order to avoid excessive delays when requests for full text are placed into the teletypewriters, a special control switch wired directly to the buffer/controller in the augmented-catalog console will allow fast-action response to commands such as requests for page-turning.

THE CENTRAL STATION

The central-station block in Fig. 12 contains a collection of the 10,000 microfilmed documents being included in the Intrex experiments and, upon request, a means for transmitting the image of any page of this collection over the transmission subsystem to a user's terminal. The requirements we have placed on this station are:

1. Access time to any document shall be ten seconds, maximum, with five seconds preferred. Access time to any page in a retrieved document shall be one second or less.

2. Transmitted image quality shall be the image quality obtainable with a scanner system with the following characteristics: on-axis limiting resolution of 1200 cycles per page, a ramp scan waveform which is within one percent of linearity, signal-to-noise ratio of 40 to 1, and linear rendition of at least six gray levels.

3. Raster-scan format shall be variable up to 2,000 lines, and raster dimension shall be variable in the horizontal and vertical dimensions. Furthermore, the scan time shall be variable from 0.5 second, minimum.

4. An ability to have all retrieval, scanning (including variations in scanning parameters), and transmission controlled from a remote computer.

5. A capability to provide appropriate control signals for the transmission network and the receiver terminals.

The scanner is being developed in our laboratory and is essentially completed; it is described in detail in our 15 September 1967 Semiannual Activity Report. The microfilm storage-and-retrieval unit is being procured from the Nuclear Research Instrument Division of the Houston-Fearless Corporation. The unit is their CARD retriever, modified optically and mechanically to accommodate our scanner assembly. The additional electronic circuitry to complete the station is under development in our laboratory.

Details of the Central Station being designed are illustrated in Fig. 13. The microfilmed images of the documents are stored on microfiche cards in a format

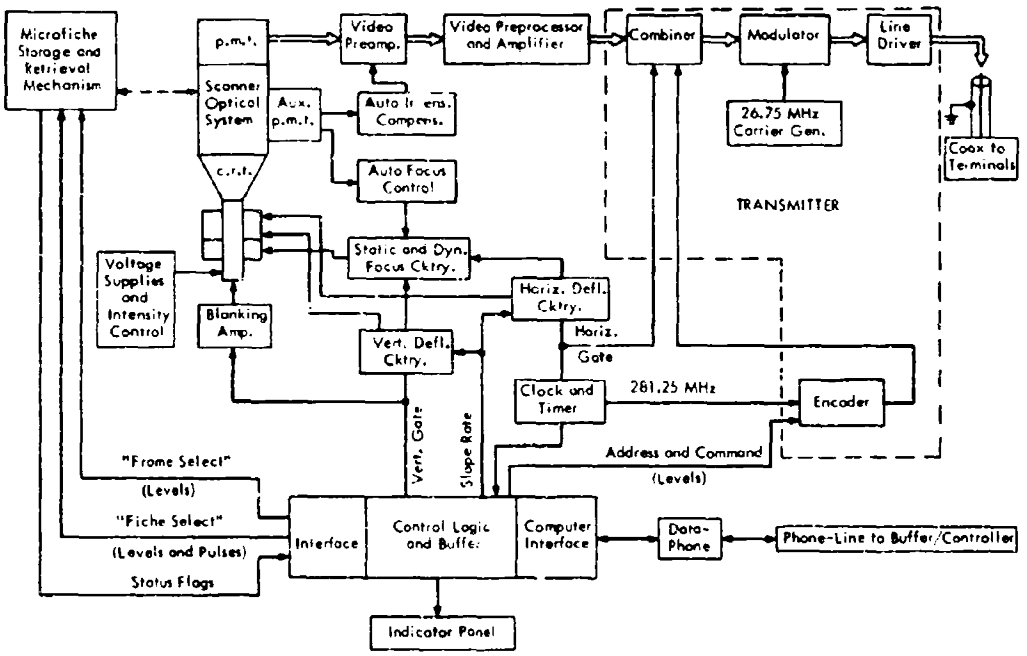


Fig. 13 Text-Access Central Station

that adheres to the COSATI standards; each fiche contains a maximum of 60 frames and each frame contains the image of one page. The frames are arranged in five rows, designated A to E, and each row contains twelve frames.

The choice of microfiche as the storage form was based on the fact that the only suitable retrieval equipment available is designed for such a form. It may well be that future text-access systems, incorporating a large number of documents, will be based on roll film or, even more likely, on film chips as the storage medium.

A metal clip is attached to the upper edge of each microfiche card and the clip is notch-coded to identify the card uniquely. An identifying number is assigned each card sequentially during preparation of the microfiche. Thus, any frame in the entire collection has a unique identification, or access number; for example, the access number 125B7 refers to the seventh frame of the second row of card No. 125.

The cards are held in a radial arrangement with their metal clips outward around a rotary tray that is part of the Houston-Fearless CARD unit. Upon receipt of an access number from the control unit, the tray rotates until the card with the corresponding notched clip is detected. The card is then placed in a holder and moved laterally until the appropriate frame is centered on the optical axis of the scanner. A signal to the vertical sweep gate of the scanner then initiates a scan and the resultant video signal flows to the transmission network. If the next requested page is on the same card, the CARD unit will move it into scanning position within one second of receipt of the new frame number. If a page on a different card is requested, the card just scanned will be replaced in the rotary tray and a new one will be retrieved. Return-to-tray is accomplished within a five-second interval.

The flying-spot scanner and its operation have been described in previous reports. This unit will remain essentially as reported, although some improvements and refinements may be added as needed. These may possibly include: automatic beam-level control; automatic spot-intensity compensator (video-gain control); automatic-focus control; and video-signal conditioning (for example, high-frequency peaking and phosphor-decay compensation). We plan to run the horizontal sweep of the scanner continuously, but the vertical sweep and unblank functions will be initiated by signals from the buffer/controller of the augmented-catalog console. The latter also sends to the coder addresses and other commands for the terminals, to be transmitted with the video signal. The buffer/controller unit is, in turn, connected to the central time-sharing computer which receives requests for access.

THE TRANSMISSION SUBSYSTEM

The transmission subsystem provides wideband video channels to link the central station with the user terminals through unidirectional coaxial cables connected in a tree network. The information to be transmitted over the cable network consists of the analog video signal, synchronization for this signal, and various digital signals such as the commands which control user-terminal operation, information for the identification of the film-strip outputs of microfilm-facsimile terminals, and an address to direct each transmission to the proper user terminal.

Signal-Design Considerations

The analog video signal has the greatest bandwidth requirement of the above signals because of the high-resolution requirements of the text-access system. Pulse-code modulation, although it has the advantages of digital transmission, would result in excessive system complexity and bandwidth requirements. Therefore, conventional analog transmission of the video signal will be used in the experimental system. The other signals transmitted are, however, well suited to pulse transmission over a channel of the bandwidth required for the video signal. A composite signal containing both analog and digital components will therefore be used.

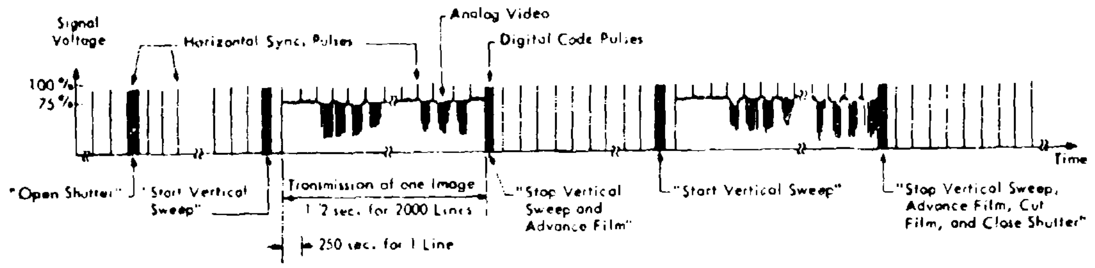
A composite signal has been designed for transmission over a single wide-band analog channel; analog and pulse signals are distinguished by their relative amplitudes and timing. Pulses have the full amplitude of the channel signal, while the maximum amplitude of the video signal, which corresponds to black level in the original image, is 75 percent of full channel signal. A typical composite waveform, corresponding to the transmission of two frames to a microfilm facsimile output terminal, is shown in Fig. 14.

The horizontal sync pulses, 25 microseconds in duration at a repetition rate of 4 KHz for the fastest anticipated scan rate, are transmitted continuosuly over the cable network. Figure 14 shows that the time between these pulses may be occupied by either no signal, or an analog video signal corresponding to one line of an image, or a sequence of pulses representing a digital word, to be discussed below. The horizontal sync pulses provide for the following functions: synchronization of the horizontal sweep circuitry at the receiver terminal; synchronization of receiver digital circuits; and automatic gain control in the receivers.

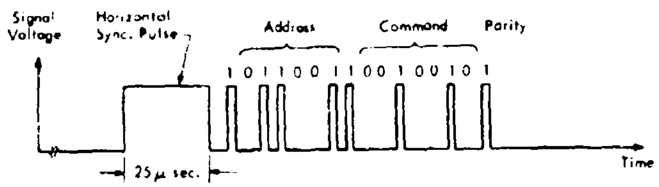
Each digital code word is represented as a temporal binary sequence in which a pulse presence corresponds to a logical one and pulse absence corresponds to a logical zero, as shown in Fig. 14. Every word is composed of 16 bits--a six-bit address, followed by a seven-bit command and various framing and parity check bits. A different address code is assigned to each user terminal; thus, commands can be sent to one terminal while the others remain idle. The standard ASCII seven-bit code has been chosen for the command so that a full character set is available for the identification of the film-strip outputs of the microfilm-facsimile terminal.

OPERATION OF THE TRANSMISSION SUBSYSTEM

Figure 15 shows the major components of the transmission subsystem consisting of the transmitter, the channel, and a typical receiver. Upon initiation of a transmission cycle by the augmented-catalog buffer/controller, the address code of



(a) A Typical Composite Signal: The Transmission of Two Images To a Microfilm Facsimile Terminal.



(b) Expanded Time Scale Graph Showing a Typical Address-Command Waveforms.

Fig. 14 Waveforms of the Transmitted Signal

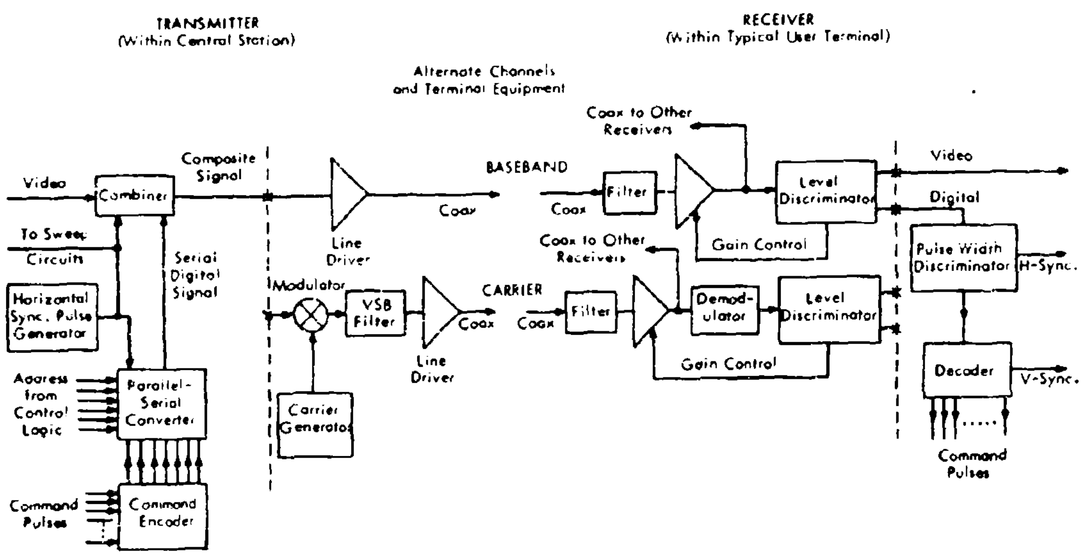


Fig. 15 The Transmission Subsystem

the particular user terminal which is to receive an image, or a set of images, is stored in a shift register within the text-access central-station control logic. At various times during the transmission cycle, commands for the user terminals are generated in different parts of the central station. Typical commands are: "start vertical sweep", "advance film" and "erase". The command encoder and parallel-serial converter shown in Fig. 15 transform these commands and the stored address code into a serial binary sequence of pulses and spaces. An example of a sequence is shown in Fig. 14a. Synchronizing circuitry ensures that this sequence always begins at a fixed time after the trailing edge of the horizontal sync pulse. The combiner superimposes the horizontal sync and digital pulses on the video signal in their proper relative amplitudes, resulting in a signal such as that shown in Fig. 14a.

The composite signal requires a single channel of 4 MHz bandwidth. Two alternatives for the text-access system are indicated in Fig. 15, namely, carrier or baseband transmission over the coaxial network. When low-frequency noise susceptibility, signal attenuation, linear distortion, and system complexity are considered, each method has its own advantages. Because of uncertainty in several system parameters involved, we have decided to determine experimentally which method is better suited for this application.

The filter in each receiver is intended to compensate for linear distortion in the cable and to exclude unwanted frequencies. An automatic gain control in each receiver will compensate for slow variations of transmission loss to provide proper contrast in the received image and proper discrimination of pulse amplitudes. A level discriminator distinguishes pulses from the analog video signal. A pulse-width discriminator distinguishes the wide horizontal sync pulses from the digital pulses and accepts only digital pulses which are in proper phase with the trailing edge of the sync pulse. The decoder of a particular receiver recognizes only those digital words which have the proper address. The command character of such a word is decoded so that a pulse appears on the appropriate decoder output. Video reception is activated upon the recognition of an appropriate command.

THE USERS' TERMINALS

The usefulness of future operational text-access systems will depend to a great extent on the capabilities of the users' terminals. Low cost for the terminals will obviously be of paramount importance. In addition, we are presently facing, in the realm of terminals, severe technological limitations, especially in the area of transient displays. We note, however, substantial research and development activities in the industrial sector which should, if successful, contribute markedly to text-access-terminal advancements. In light of these external pending developments we plan, for the present, to employ a minimum number of terminals. We shall have one terminal for each major type of output device, that is, a terminal for transient display of text, another for film copy, and a third for experimenting with new devices as they come along. We also expect to have a separate terminal for making paper copy.

We are favoring the transient-type terminal, since it approaches most closely the capability of providing immediate text access and is potentially the least expensive to operate. Unfortunately, no fully satisfactory device is available for our purpose; however, the Tektronix eleven-inch storage tube does afford limited capabilities and it will be incorporated into one terminal. The second class of terminal, the film terminal, will provide text with higher resolution than the Tektronix display. The film terminal will also enable us to test the acceptability of film as a primary hard-copy output; that is, we shall be able to decide if film is an acceptable substitute for the more expensive paper copy.

Storage-Tube Terminal

The storage-tube terminal, diagrammed in Fig. 16, makes use of the Tektronix Type 611, eleven-inch storage display unit. An evaluation of an experimental engineering model of the Tektronix direct-view storage-tube display was reported in the preceding Intrex Semiannual Activity Report. We have concluded that the resolution and brightness of this display are adequate for the reader who wishes to make a preliminary examination of text in order to verify its relevance to his requirements. Resolution may be marginal, however, for perception of poor-quality print or small symbols and characters. A possible way to overcome the resolution limitation is to display an enlarged version of a portion of a page of text.

Operation of the storage-tube terminal is straightforward. Upon receipt of an ERASE command from the demodulator, any image appearing on the screen is erased. One-half second later the BEGIN SWEEP command will be received, followed by the video signal for one page of text. After being written on the screen, the text remains on the tube face until the next ERASE command. Because the writing speed of the Tektronix Type 611 display is rather slow, the scanning of a frame of the original microfilm and the transmission of the corresponding signal must be extended from our desired one-half second writing time to four seconds.

Microfilm-Facsimile Terminal

The microfilm-facsimile terminal, drawn in Fig. 17, consists of a high-resolution cathode-ray tube with its associated sweep and focus circuitry, an automatic camera-processor, and control logic required to operate the terminal. On command from the central station the microfilm-facsimile terminal will reconstitute on the face of a high-resolution cathode-ray tube the image of a full page of text from a video signal of 4.5-MHz bandwidth transmitted from the central station. The automatic-camera and film-processor unit will record on 35-mm film the image of the text displayed on the cathode-ray tube and deliver to the user a fully processed strip of film in a convenient form for viewing in a microfilm reader.

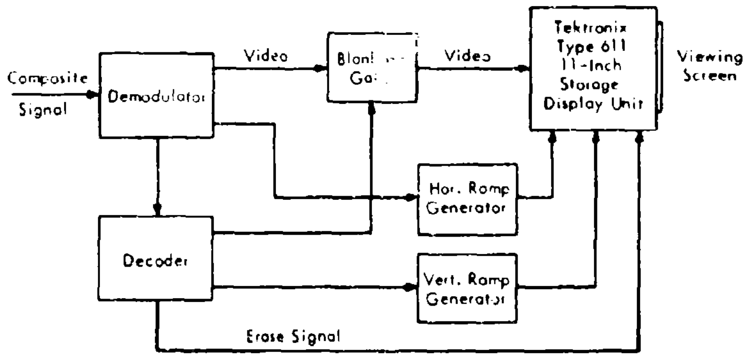


Fig. 16 Storage-Tube Terminal

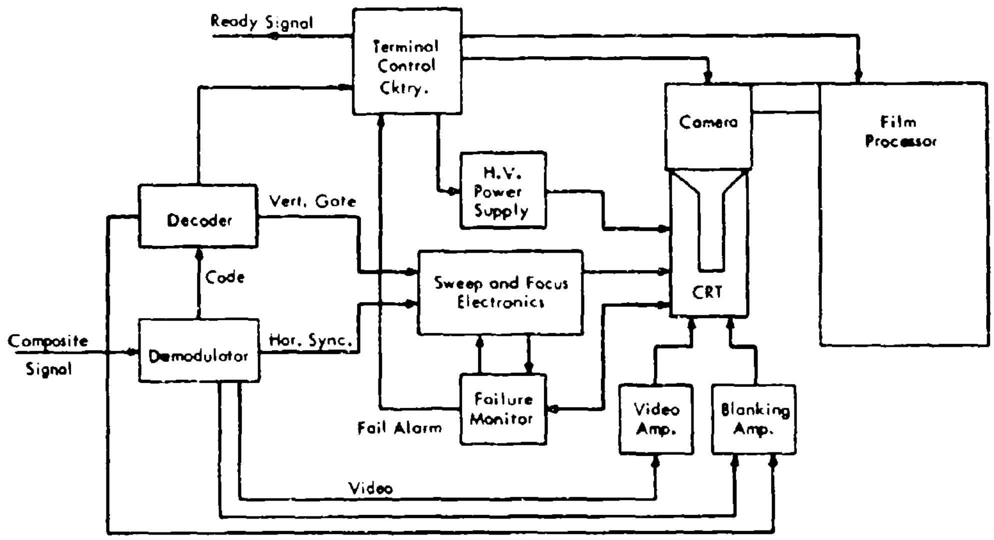


Fig. 17 The 35-mm Film Copy Terminal

The operation and configuration of the high-resolution cathode-ray tube with its associated sweep and focus circuitry is described in the Intrex Semiannual Activity Report dated 15 March 1967. The operational requirements for the camera and film processor are as follows:

1. The automatic-camera and film-processor unit shall record on 35-mm film the image of a full page of text which is obtained in a single scan and displayed on the screen of a high-resolution cathode-ray tube. It shall also deliver to the user a fully processed strip of film in a convenient form for use in a microfilm viewer.

2. Each strip of film will contain a minimum of one, and a maximum of ten, adjacent images.

3. The maximum combined length of unexposed leader and trailer on each film strip shall be five inches.

4. The film transport of the camera and processor shall handle unperforated 35-mm film.

5. The microfilm-facsimile terminal shall not require an attendant for normal operations and the camera and processor shall not require routine maintenance, other than the loading of film and chemicals, more than once per week.

6. The camera-and-processor unit shall be designed for operation by electrical-control signals.

7. In view of the experimental nature of the terminal, the camera-and-processor unit shall be designed with emphasis on flexibility; that is, it shall be possible to change the type of film, the size of the image on the film, the type of chemicals utilized, or the lens, without major equipment alterations.

No camera-and-processor unit that satisfactorily meets all the above requirements was found to be commercially available. Our plan therefore is to purchase a microfilm camera and a separate, leaderless-type, film processor and to merge the two units into a camera-and-processor unit. A Kodak MCD-II film unit and a GAF Transflo Type 1206 Processor have been procured, and detailed drawings of the required modifications are nearing completion.

Figure 18 illustrates the modifications in the Kodak MCD-II film unit which are required in order to transport short strips of 35-mm film from the camera to the film processor automatically. The take-up reel has been replaced by a cutter mechanism that will, upon application of the appropriate electrical signals, sever the exposed film strip and transport it to the processor.

Each exposed strip of film contains from one to ten contiguous images and a four-inch leader. At the end of a series of exposures, or after ten adjacent frames have been exposed, the camera is commanded to advance the film until the last exposed frame has cleared the film cutter. The solenoid-operated film-cutter then severs the film.

Belt-driven Bendix electric clutches are employed to ensure that the film-transport rollers are synchronized with the camera during frame advance and with the processor, as the film is fed to that machine. During frame advance, clutch A in Fig. 18 is engaged. This action connects the existing camera drive to the new

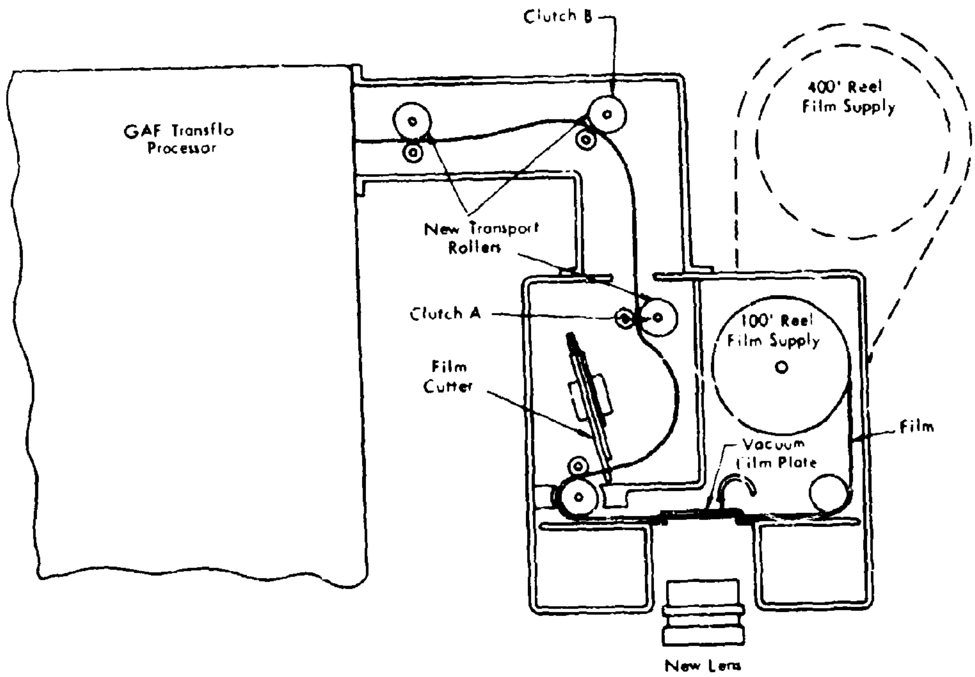


Fig. 18 Modified MCD-2 Camera Head

transport rollers. Similarly, during the feeding of the processor, clutch B is engaged, thereby connecting the processor drive to the transport rollers.

As shown in Fig. 18, the modified camera assembly is attached directly to the GAF automatic processor. This processor utilizes a horizontal straight-line film transport that is self-threading and will, with minor modifications, accept short strips of 35-mm film. It should be noted that the GAF Transflo film processor was designed for films of 12 inches maximum width, and for a much heavier volume of processing than we anticipate. It appeared after survey of available film processors that the GAF Transflo machine comes closest to meeting our needs, at least on a temporary basis.

It should be evident that success of the microfilm-facsimile terminal depends on the availability of convenient microfilm viewers adapted to handle the short film strips that will emerge from the camera-processor unit. Again, no viewer has been found which exactly meets our requirements. We must, therefore, devise a suitable method of viewing film strips by making modifications to a viewer now in our laboratory, an IBM 9922 document viewer. We plan to place the film strips into simple cardboard mountings which can be inserted into the viewer.

MICROFILMING FOR THE EXPERIMENTAL TEXT-ACCESS SYSTEM

The external dimensions of the microfiches to be used in the experimental text-access system are nominally 6 in. x 4 in. The internal arrangement of the images on the fiches and the space allocated to each page will conform essentially to the National Microfilm Association, COSATI, and the impending USA Standards Specifications for fiches.

The resolution of the film images will be between 120 and 150 lines per millimeter, that is, between 1700 and 2100 cycles per frame height. The images will be negative in tonality, and the clear areas representing the actual text will be kept at minimum density. The background density of the images will be adjusted to achieve maximum contrast compatible with keeping the lines of the text clear and open. Since the fiches are not intended to be circulated or used manually, no macroscopic title will be needed, but at the top of each fiche there will be an identifying number.

Two characteristics of the text-access system impose restrictions on the placing of the page images relative to the microfiche grid. One characteristic is that the resolution capability of the scanning system is below that of the microfilm. In order to minimize this resolution degradation the scanning raster should coincide as closely as possible with the image of the text area. Another characteristic, however, is that the scanning raster and the position of the fiche in the scanner mechanism are preset, so that the user has no control over the positioning of the page image as he views it (a control one does have in manual viewers). If the image is not congruent with the scanning raster, part of it will be lost. Therefore, page images must be accurately centered in a regular grid pattern which, in turn, is accurately positioned with respect to the edges of the microfiche.[3]

---

[3] Unsatisfactory alternatives are: (1) Provision for user-controlled fiche position adjustment--this approach requires retransmission and is therefore wasteful as well as inconvenient; (2) Increasing the size of the scanning raster and accepting the resultant lower resolution as well as the possible unaesthetic lopsidedness of the image; (3) Incorporation of special markers in the image frame and detecting them for automatic repositioning--a scheme that may be quite satisfactory in the present system, but would be impractical for future microfilmed material from external sources; (4) Automatic detection of the edges of the text image itself, which requires a rather sophisticated device, if it could be made reliable at all.

This requirement, as well as the one needed for minimal reduction ratio, calls for two important revisions in standard microfilming procedure: when material is filmed from bound volumes, each page must be filmed separately, rather than in a double-page spread format, as is customary; positioning of the page during filming, and of the film frames in the camera and in subsequent steps in the preparation of the microfiche, must be carefully controlled. In order to facilitate the first revision, as well as, in part, the second one, a special book cradle is being designed and constructed, as described in the following pharagraph.

The book cradle which is in a process of construction will consist of a book-holding device which is glass-topped and V-shaped. The V-shaped book holder is in turn supported by a carriage which allows oscillation of the entire book holder so as to present the left- and right-hand pages alternately to the camera lens. The purpose of the V shape of the cradle is to secure the gutter of the book in the crotch of the V to prevent the pages from creeping sideways or from presenting the text in an askew position. As the oscillating carriage is moved from left to right, and back, the V of the book holder seesaws so as to present the relevant leg of the V squarely to the lens and parallel to the film plane. For the time being, only the oscillating-carriage part has been constructed and will be used with a flat, rather than the V-shaped, book holder. Both are shown mounted on the camera table in the photograph of Fig. 19.

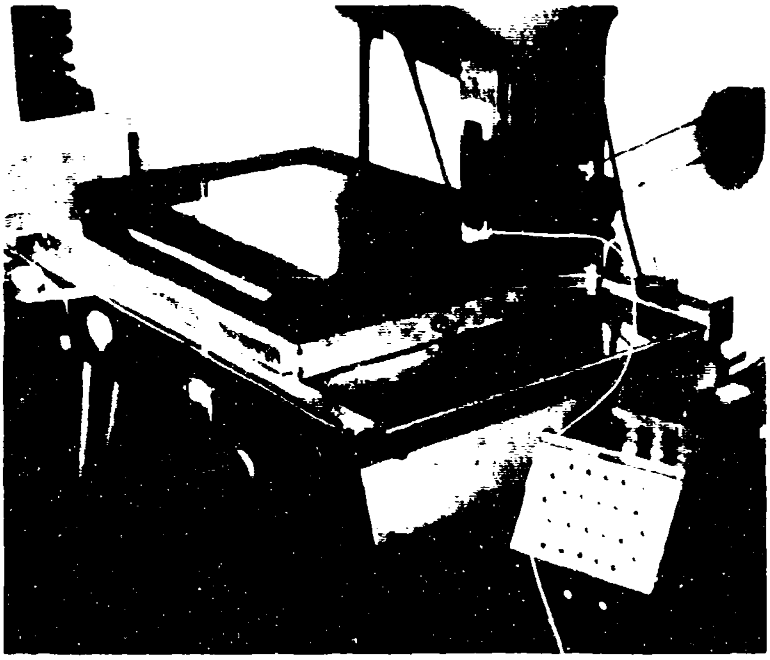


Fig. 19 Camera Table with Oscillating Carriage, Flat Book Holder and Frame Counter

The preparation of the microfiches includes the following steps: The material, after submission by the librarian, will be recorded on 16-mm microfilm on a planetary camera with several special features. A commercial kit has been installed which renders the page-to-page spacing on the film more constant than is the case in normal microfilming. The effect is to ensure the proper placement of the images on the final microfiche. The pages are placed on the roll film in a pattern which facilitates subsequent separation and mounting of short strips of the film, so that a stripped-up microfiche master can be assembled. Essentially this process consists of the recording of twelve microfilm images, an interval of clear film, the recording of a further twelve-page group, and so on. In order that the camera operator not be required to keep track of the proper sequencing of the images, a special counter was designed which, by virtue of visual and audible signals, keeps track of the rows and intervals and of the completion point of each fiche. This counter is tied to the foot switch which activates the camera shutter, and is shown at the right side of Fig. 19.

The roll film is then processed in a large, continuous film processor with above-average processing controls. The processing produces film of archival permanence. Before the 16-mm roll-film master is cut and stripped, a duplicate film is made. The duplicate has two functions. It serves as a spare master in the event of damage of any kind to the original film and as an intermediate for the generation of Xerox copies of the text for the catalogers who prepare the data base. The duplicate film utilizes a silver-reversal film which yields a duplicate negative from the original (master) negative. The duplicate is then fed to a Xerox electrostatic continuous enlarger and the resultant paper copies are channeled to the catalogers.

Subsequently, the original, which is handled in 100-foot rolls, is placed on a combination cutter and stage. The film is cut into suitable strips and each cut is accomplished by an automatic hole punch close to and on both sides of the separation. The resultant strips have two holes, one near each end, which fit over two prongs on the strip-up station. The fiche master is completed with the arrangement, in this manner, of five film strips and a top strip bearing the typewritten identification number on a translucent material. A heavy-weight adhesive tape with prepunched holes fits vertically across the prongs and film strips on each side so that the whole array may be lifted off for a permanent, rigid arrangement. The masters are then stored in glassine envelopes.

The actual copy to be placed in the storage and retrieval unit is prepared by contact-printing the stripped-up master to a diazo duplicate fiche. A special jig may have to be utilized here to maintain the required position accuracy. Finally, a notch-coded metal strip is attached to the fiche, again in accurate alignment. The fiche is then ready to be placed in the modified Houston-Fearless storage and retrieval unit of the remote text-access system.

III. PROJECT INTREX STAFF

A. PROJECT OFFICE

Professor Carl F. J. Overhage, Director
Mr. Joseph J. Beard
Mr. Charles H. Stevens

B. ELECTRONIC SYSTEMS LABORATORY

Professor J. Francis Reintjes, Director
Mr. John E. Ward, Deputy Director
Mr. Alan R. Benenfeld
Mrs. Sonja K. Escudier
Dr. Uri F. Gronemann
Mrs. Elizabeth J. Gurley
Dr. Donald R. Haring
Mr. Charles E. Hurlburt
Mr. James E. Kehr
Mr. Donald R. Knudson
Mr. Peter Kugel
Mr. Robert L. Kusik
Mrs. Sondra F. Lage
Miss Lucy T. Lee
Mr. Richard S. Marcus
Mr. Michael K. Molnar
Miss Sonia P. Niessen
Professor James K. Roberge
Miss Jane E. Rust
Mr. Peter R. Scott
Professor Alfred K. Susskind
Mr. Stephen N. Teicher
Mr. Herman F. Vandevenne

C. ENGINEERING LIBRARY

Miss Rebecca L. Taggart, Head
Miss Margaret W. Artinian
Miss Barbara C. Darling
Mrs. Jean Field
Mr. Jeffrey L. Gardner
Mr. James M. Kyed
Mrs. Suanne W. Muehlner
Mrs. Alice M. Robrish
Mrs. Colleen M. Scholz
Mrs. Ines Siscoe

IV. PUBLICATIONS

CONFERENCE PAPERS

Benenfeld, A. R., "Data Encoding for the Project Intrex Augmented-Catalog Experiments." Presented at the ADI Convention, User Discussion Group on Emerging Machine-Readable Tape Services, New York, October 25, 1967.

Stevens, C. H., "Project Intrex - Plans and Progress." New York Chapter, Special Librarians Association, New York, February 6, 1968.

REPORTS

Haring, D. R., and Roberge, J. K., "The Augmented-Catalog Console for Project Intrex," ESL-TM-323, October, 1967.